GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 27, 2003, 06:45:13 ; Search time 23 Seconds
(without alignments)
372.124 Million cell updates/sec

Title: US-09-965-594-1
Perfect score: 953
Sequence: 1 MAPITAYAQQTRGLLGCIIT.....GVAKAVDFIPVESLETTMRS 182

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 99%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	943	99.0	3011	1	POLG_HCV1	P26664 h genome po
2	927	97.3	3011	1	POLG_HCVH	P27958 h genome po
3	892	93.6	3010	1	POLG_HCVTW	P29846 h genome po
4	891	93.5	3010	1	POLG_HCVBK	P26663 h genome po
5	887	93.1	3010	1	POLG_HCVJA	P26662 h genome po
6	884	92.8	3010	1	POLG_HCVJT	Q00269 h genome po
7	714	74.9	3033	1	POLG_HCVJ8	P26661 h genome po
8	712	74.7	3033	1	POLG_HCVJ6	P26660 h genome po
9	87	9.1	209	1	PAAD_PSEAE	Q9hx08 pseudomonas
10	84	8.8	321	1	HHOA_ARATH	Q9sel7 arabidopsis
11	82	8.6	452	1	AAMP_HUMAN	Q13685 homo sapien
12	82	8.6	485	1	Y136_TREPA	O83172 treponema p
13	80.5	8.4	437	1	DEG1_ARATH	O22609 arabidopsis
14	75	7.9	253	1	CAC3_BOVIN	P05805 bos taurus
15	74.5	7.8	415	1	ZP3_RABIT	P48833 oryctolagus
16	74.5	7.8	776	1	HYPF_AZOVI	P40596 azotobacter
17	74.5	7.8	911	1	TB11_NEIMB	Q09056 neisseria m
18	74	7.8	326	1	PANE_RHILO	Q987n5 rhizobium l
19	73.5	7.7	263	1	GRAK_MOUSE	O35205 mus musculu
20	73	7.7	730	1	HELS_METMA	Q8pzr7 methanosarc
21	72.5	7.6	257	1	GRAM_HUMAN	P51124 homo sapien
22	72	7.6	627	1	SAD1_MOUSE	Q60710 mus musculu
23	72	7.6	1527	1	CA1H_MOUSE	P39061 mus musculu
24	72	7.6	2663	1	CENE_HUMAN	Q02224 homo sapien
25	72	7.6	3491	1	ERY1_SACER	Q03131 saccharopol
26	71.5	7.5	248	1	GRAD_MOUSE	P11033 mus musculu
27	71.5	7.5	323	1	VPRT_SMRVH	P21407 squirrel mo
28	71	7.5	219	1	SPR1_IPOBA	P14715 ipomoea bat
29	71	7.5	336	1	UL16_EBV	P03221 epstein-bar
30	71	7.5	529	1	PGL2_RALSO	P20041 ralstonia s
31	71	7.5	1180	1	ITA1_RAT	P18614 rattus norv
32	70.5	7.4	264	1	CTRL_HUMAN	P40313 homo sapien
33	70.5	7.4	659	1	VST2_HEVME	Q03500 hepatitis e
34	70.5	7.4	743	1	TFE3_HUMAN	P19532 homo sapien
35	70	7.3	280	1	VIE3_HCMVT	P06434 human cytom
36	70	7.3	410	1	VIE2_HCMVT	P06435 human cytom
37	70	7.3	478	1	MM03_RABIT	P28863 oryctolagus
38	70	7.3	730	1	HELS_METAC	Q8tl39 methanosarc
39	69.5	7.3	915	1	TBP1_NEIGO	Q01996 neisseria g
40	69	7.2	355	1	CWG2_SCHPO	P32434 schizosacch
41	69	7.2	637	1	DNAK_BRUME	Q8ye76 brucella me
42	69	7.2	1136	1	C4BA_BACTI	P05519 bacillus th
43	69	7.2	1180	1	C4AA_BACTI	P16480 bacillus th
44	69	7.2	4391	1	PGBM_HUMAN	P98160 homo sapien
45	68.5	7.2	255	1	CATG_HUMAN	P08311 homo sapien

ALIGNMENTS

RESULT 1
POLG_HCV1
ID POLG_HCV1 STANDARD; PRT; 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA Bradley D.W., Kuo G., Houghton M.;
RT "Genetic organization and diversity of the hepatitis C virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC ---

CC ---
DR EMBL; M62321; AAA45676.1; -.
DR PIR; A39166; GNWVC3.
DR PDB; 1A1V; 16-FEB-99.
DR PDB; 1HEI; 25-NOV-98.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.

```
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3011       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    476    476       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2041   2041       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2077   2077       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2240   2240       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2364   2364       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2789   2789       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3011 AA;  327197 MW;  65F8C9447FCE5AF9 CRC64;

  Query Match           99.0%;  Score 943;  DB 1;  Length 3011;
  Best Local Similarity  98.4%;  Pred. No. 3.4e-82;
  Matches  179;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :|||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 LAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db       1086 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              |||||||||'||||||||||||||||||||||||||||||||||||||||||||:|||||
Db       1146 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 1205

Qy        181 RS 182
              ||
Db       1206 RS 1207


RESULT 2
POLG_HCVH
ID   POLG_HCVH      STANDARD;      PRT;  3011 AA.
AC   P27958;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate H) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11108;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92052256; PubMed=1658800;
RA   Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA   Prince A.M.;
RT   "Genomic structure of the human prototype strain H of hepatitis C
RT   virus: comparison with American and Japanese isolates.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX   MEDLINE=97331322; PubMed=9187654;
RA   Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT   "Structure of the hepatitis C virus RNA helicase domain.";
RL   Nat. Struct. Biol. 4:463-467(1997).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX   MEDLINE=98154321; PubMed=9493270;
RA   Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA   Murcko M.A., Lin C., Caron P.R.;
RT   "Hepatitis C virus NS3 RNA helicase domain with a bound
RT   oligonucleotide: the crystal structure provides insights into the mode
RT   of unwinding.";
RL   Structure 6:89-100(1998).
CC   -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC   -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC       NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC   -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC       ACTIVATION OF NS3.
CC   -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC   -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC       ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC       AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
```

```
CC   -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC       PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC   -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC   -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M67463; AAA45534.1; -.
DR   PIR; A36814; GNWVCH.
DR   PDB; 1HEI; 25-NOV-98.
DR   PDB; 1A1V; 16-FEB-99.
DR   PDB; 1A1R; 17-JUN-98.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   TRANSFAC; T04155; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    191       CAPSID PROTEIN C.
FT   CHAIN       192    383       ENVELOPE GLYCOPROTEIN E1.
FT   CHAIN       384    746       ENVELOPE GLYCOPROTEIN E2.
FT   CHAIN       747    809       PROTEIN P7.
FT   CHAIN       810   1026       NONSTRUCTURAL PROTEIN NS2.
FT   CHAIN      1027   1657       PROTEASE/HELICASE NS3.
FT   CHAIN      1658   1711       NONSTRUCTURAL PROTEIN NS4A.
FT   CHAIN      1712   1972       NONSTRUCTURAL PROTEIN NS4B.
FT   CHAIN      1973   2420       NONSTRUCTURAL PROTEIN NS5A.
FT   CHAIN      2421   3011       NONSTRUCTURAL PROTEIN NS5B.
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    476    476       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   STRAND     1224   1226
FT   TURN       1232   1233
FT   TURN       1236   1238
FT   HELIX      1239   1246
FT   TURN       1247   1248
FT   STRAND     1251   1255
FT   HELIX      1258   1271
FT   TURN       1272   1272
FT   STRAND     1277   1280
FT   TURN       1281   1282
FT   STRAND     1283   1285
FT   STRAND     1291   1295
FT   HELIX      1296   1301
FT   TURN       1302   1303
FT   STRAND     1312   1316
FT   TURN       1317   1319
FT   HELIX      1323   1335
FT   TURN       1336   1340
FT   STRAND     1343   1347
FT   TURN       1352   1353
FT   TURN       1360   1361
FT   STRAND     1362   1366
FT   STRAND     1368   1368
FT   STRAND     1373   1375
FT   TURN       1376   1377
FT   STRAND     1378   1380
FT   HELIX      1382   1385
FT   STRAND     1389   1393
FT   HELIX      1397   1409
FT   TURN       1410   1411
FT   STRAND     1414   1417
FT   TURN       1419   1420
FT   STRAND     1432   1436
FT   TURN       1438   1439
FT   STRAND     1450   1453
FT   STRAND     1456   1463
FT   STRAND     1471   1478
FT   STRAND     1480   1480
FT   HELIX      1481   1488
FT   TURN       1489   1490
FT   STRAND     1497   1501
FT   STRAND     1507   1507
FT   STRAND     1511   1511
FT   HELIX      1514   1527
FT   HELIX      1532   1544
FT   STRAND     1550   1550
FT   HELIX      1555   1564
FT   HELIX      1570   1578
FT   TURN       1579   1580
FT   HELIX      1584   1597
FT   TURN       1598   1598
FT   HELIX      1606   1611
FT   TURN       1614   1618
FT   STRAND     1622   1623
FT   STRAND     1627   1627
FT   STRAND     1635   1636
FT   HELIX      1640   1652
SQ   SEQUENCE   3011 AA;  327142 MW;  772CBB29CCD94753 CRC64;

  Query Match           97.3%;  Score 927;  DB 1;  Length 3011;
  Best Local Similarity  96.2%;  Pred. No. 1.2e-80;
```

```
  Matches  175;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :|||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db       1026 LAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTATQTFLATCINGVCWTVYHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              |||||||||||||| ||||| :||||||||||||||||||||||||||||||||||||||
Db       1086 GTRTIASPKGPVIQTYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              ||||||||||||||||||||||||||| |||||:|||||||||||||||||||| |||||
Db       1146 GDSRGSLLSPRPISYLKGSSGGPLLCPTGHAVGLFRAAVCTRGVAKAVDFIPVENLETTM 1205

Qy        181 RS 182
              ||
Db       1206 RS 1207


RESULT 3
POLG_HCVTW
ID   POLG_HCVTW     STANDARD;      PRT;  3010 AA.
AC   P29846;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate Taiwan) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31645;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92230206; PubMed=1314449;
RA   Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
RT   "The Taiwanese hepatitis C virus genome: sequence determination and
RT   mapping the 5' termini of viral genomic and antigenomic RNA.";
RL   Virology 188:102-113(1992).
CC   -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC       HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC       NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       (RNA)(N).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M84754; -; NOT_ANNOTATED_CDS.
DR   PIR; A40244; GNWVTW.
DR   PDB; 1N64; 25-FEB-03.
DR   PDB; 1NS3; 08-APR-98.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CORE PROTEIN (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    233    233       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2041   2041       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2077   2077       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2240   2240       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2529   2529       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2788   2788       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3010 AA;  327047 MW;  AAD267D55CDFE215 CRC64;

  Query Match             93.6%;  Score 892;  DB 1;  Length 3010;
  Best Local Similarity   90.1%;  Pred. No. 2.7e-77;
  Matches  164;  Conservative   12;  Mismatches    6;  Indels    0;  Gaps    0;
```

```
Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :|||||:|||||||:| :|||||||||||||:|||||||:|||| :|||:|||||| ||||:
Db       1026 LAPITAYAQQTRGLFGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVYHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              |:|.:| ||||: ||.|:|||:|||| ||| :|||:|||||:|||:|||||||:|||||:
Db       1086 GSKTLAGPKGPITQMYTNVDQDLVGWHAPQGARSLIPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              |||||:|||||:|||||||:|||||||:|| |:|||:|||||||:|||||:|||||:||| 
Db       1146 GDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFVPVESMETTM 1205

Qy        181 RS 182
              || 
Db       1206 RS 1207
```

```
RESULT 4
POLG_HCVBK
ID   POLG_HCVBK     STANDARD;      PRT;  3010 AA.
AC   P26663;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate BK) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11105;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91140698; PubMed=1847440;
RA   Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA   Onishi E., Andoh T., Yoshida I., Okayama H.;
RT   "Structure and organization of the hepatitis C virus genome isolated
RT   from human carriers.";
RL   J. Virol. 65:1105-1113(1991).
RN   [2]
RP   SEQUENCE OF 1487-1500.
RX   MEDLINE=96235224; PubMed=8647104;
RA   Borowski P., Heiland M., Oehlmann K., Becker B., Korneteky L.;
RT   "Non-structural protein 3 of hepatitis C virus inhibits
RT   phosphorylation mediated by cAMP-dependent protein kinase.";
RL   Eur. J. Biochem. 237:611-618(1996).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX   MEDLINE=97015088; PubMed=8861916;
RA   Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA   Moomaw E.W., Adachi T., Hostomska Z.;
RT   "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT   trypsin-like fold and a structural zinc binding site.";
RL   Cell 87:331-342(1996).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX   MEDLINE=98227846; PubMed=9568891;
RA   Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA   Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RT   "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RT   virus: a 2.2-A resolution structure in a hexagonal crystal form.";
RL   Protein Sci. 7:837-847(1998).
CC   -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC       HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC       NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate + N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M58335; AAA72945.1; -.
DR   PIR; A38465; GNWVIC.
DR   PDB; 1A1Q; 25-MAR-98.
DR   PDB; 1JXP; 14-JAN-98.
DR   PDB; 1NS3; 08-APR-98.
DR   PDB; 1C2P; 15-NOV-00.
DR   PDB; 1CSJ; 08-NOV-99.
DR   PDB; 1GX5; 09-APR-02.
DR   PDB; 1GX6; 10-APR-02.
DR   PDB; 1QUV; 26-JUN-00.
DR   PDB; 8OHM; 20-APR-99.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM.
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM.
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM.
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
```

```
FT   SITE        1316   1319        DECH BOX.
FT   CARBOHYD     196    196        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     209    209        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     234    234        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     250    250        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     305    305        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     423    423        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     430    430        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     448    448        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     532    532        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     540    540        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     556    556        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     576    576        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     623    623        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     645    645        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2041   2041        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2077   2077        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2240   2240        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2529   2529        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2788   2788        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   STRAND      1031   1035
FT   HELIX       1039   1047
FT   STRAND      1050   1050
FT   STRAND      1059   1063
FT   STRAND      1068   1074
FT   TURN        1075   1076
FT   STRAND      1077   1081
FT   HELIX       1082   1085
FT   TURN        1086   1087
FT   STRAND      1090   1092
FT   TURN        1093   1094
FT   STRAND      1095   1097
FT   STRAND      1101   1103
FT   TURN        1104   1107
FT   STRAND      1108   1112
FT   STRAND      1120   1120
FT   STRAND      1122   1122
FT   STRAND      1129   1133
FT   TURN        1135   1136
FT   STRAND      1139   1144
FT   STRAND      1149   1157
FT   HELIX       1158   1161
FT   TURN        1162   1163
FT   TURN        1165   1166
FT   STRAND      1168   1171
FT   TURN        1172   1174
FT   STRAND      1175   1186
FT   TURN        1187   1188
FT   STRAND      1189   1197
FT   HELIX       1198   1202
FT   TURN        1203   1204
FT   STRAND      1680   1688
SQ   SEQUENCE   3010 AA;  327189 MW;  F8422D5ECCFDFD9C CRC64;

  Query Match           93.5%;  Score 891;  DB 1;  Length 3010;
  Best Local Similarity  89.0%;  Pred. No. 3.4e-77;
  Matches  162;  Conservative   15;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :|||||:|||||:|||||||||||||||||||:|:|:| |:|:|||:|:|||| |:|
Db       1026 LAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              |::|:|:|||| |||||||:|||||| | |:|||||||||||||||||||||||||||
Db       1086 GSKTLAAPKGPITQMYTNVDQDLVGWPKPPGARSLTPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              |||||||||:||:||:|||||:||||| ||||||||||:|||||||||:||||:.|||
Db       1146 GDSRGSLLSPRPVSYLKGSSGGPLLCPFGHAVGIFRAAVCTRGVAKAVDFVPVESMETTM 1205

Qy        181 RS 182
              ||
Db       1206 RS 1207


RESULT 5
POLG_HCVJA
ID   POLG_HCVJA     STANDARD;      PRT;  3010 AA.
AC   P26662;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate Japanese) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91088550; PubMed=2175903;
RA   Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA   Sugimura T., Shimotohno K.;
RT   "Molecular cloning of the human hepatitis C virus genome from
RT   Japanese patients with non-A, non-B hepatitis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN   [2]
RP   DISCUSSION OF SEQUENCE.
RX   MEDLINE=91192160; PubMed=1849488;
RA   Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K.,
RA   Ohkoshi S., Shimotohno K.;
RT   "Molecular structure of the Japanese hepatitis C viral genome.";
RL   FEBS Lett. 280:325-328(1991).
CC   -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC       HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC       NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D90208; BAA14233.1; -.
DR   PIR; A39253; GNWVCJ.
DR   HSSP; P26663; 1JXP.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
```

```
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT   CHAIN       730   1006       NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE    1083   1083      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1107   1107      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1165   1165      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND     1230   1237      ATP (POTENTIAL).
FT   SITE        1316   1319      DECH BOX.
FT   CARBOHYD     196    196      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     209    209      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     234    234      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     250    250      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     305    305      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     423    423      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     430    430      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     448    448      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     532    532      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     556    556      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     576    576      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     623    623      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     645    645      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2041   2041      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2077   2077      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2240   2240      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2788   2788      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3010 AA;  327017 MW;  AA993794F46DB185 CRC64;

  Query Match             93.1%;  Score 887;  DB 1;  Length 3010;
  Best Local Similarity   87.9%;  Pred. No. 8.2e-77;
  Matches  160;  Conservative   17;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :|||||:|||||:|||||||||||||||||:|||:::'| |::'|||:.||||:|||||
Db       1026 LAPITAYSQQTRGLLGCIITSLTGRDKNQVDGEVQVLSTATQSFLATCVNGVCWIVYHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              :::|:| |||| :|||||:|||||'|| |:|:|| : |||'| ||||:|||:|| ||
Db       1086 GSKTLAGPKGPITQMYTNVDQDLVGWPAPQGARSMTPCTCGSSDLYLVTRHADVVPVRRR 1145

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              |||:|||||.|||||:||||||||||:|| |||||||:|||||||:|||| |:||||:
Db       1146 GDSRGSILSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFIPVESMETTM 1205

Qy        181 RS 182
              ||
Db       1206 RS 1207


RESULT 6
POLG_HCVJT
ID   POLG_HCVJT     STANDARD;      PRT:  3010 AA.
AC   Q00269;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate HC-JT) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31642;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92295714; PubMed=1318627;
RA   Tanaka T., Kato N., Nakagawa M., Cotsuyama Y., Cho M.J.,
RA   Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT   "Molecular cloning of hepatitis C virus genome from a single Japanese
RT   carrier: sequence variation within the same individual and among
RT   infected individuals.";
RL   Virus Res. 23:39-53(1992).
CC   -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC       HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC       NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       (RNA)(N).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D11168; BAA01943.1; -.
DR   PIR; A45573; A45573.
DR   PDB; 1A1Q; 25-MAR-98.
DR   PDB; 1JXP; 14-JAN-98.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
```

```
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2041   2041       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2077   2077       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2240   2240       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2529   2529       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2788   2788       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3010 AA;  326573 MW;  94A1C77435D642B5 CRC64;

  Query Match             92.8%;  Score 884;  DB 1;  Length 3010;
  Best Local Similarity   88.5%;  Pred. No. 1.6e-76;
  Matches  161;  Conservative   15;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :|||||||||||||||||,|:|||,|||||.||||||:|||| |:||||:|:|||||||: ||
Db       1026 LAPITAYAQQTRGLLGCIVTSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              |:::|:| ||||: |||||||:|||||| || |:'||||'||||||'|||||||||||:|||
Db       1086 GSKILAGPKGPITQMYTNVDQDLVGWHAPPGARSLTPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              || |||||||||:|||||||||||||||:||||| ||||,|||||,|||:|||||:|||||
Db       1146 GDGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTM 1205

Qy        181 RS 182
              ||
Db       1206 RS 1207


RESULT 7
POLG_HCVJ8
ID   POLG_HCVJ8     STANDARD;      PRT;  3033 AA.
AC   P26661;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate HC-J8) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11115;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92230232; PubMed=1314459;
RA   Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA   Fukuda S., Tsuda F., Mishiro S.;
RT   "Full-length sequence of a hepatitis C virus genome having poor
RT   homology to reported isolates: comparative study of four distinct
RT   genotypes.";
RL   Virology 188:331-341(1992).
CC   -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC       HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC       NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       (RNA)(N).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10988; BAA01761.1; -.
DR   PIR; A40250; GNWVJ8.
DR   HSSP; P27958; 1HEI.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
```

```
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    733       NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT   CHAIN       734   1010       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1011   1629       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1630   1866       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1867   2017       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2018   3033       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE    1087   1087      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1111   1111      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1169   1169      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND     1234   1241      ATP (POTENTIAL).
FT   SITE        1320   1323      DECH BOX.
FT   CARBOHYD     196    196      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     209    209      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     233    233      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     299    299      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     305    305      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     423    423      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     430    430      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     448    448      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     477    477      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     534    534      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     542    542      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     558    558      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     578    578      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     627    627      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     649    649      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1091   1091      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2038   2038      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2359   2359      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2811   2811      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3033 AA;  330177 MW;  1A173E7E3381FD1A CRC64;

  Query Match             74.9%;  Score 714;  DB 1;  Length 3033;
  Best Local Similarity   70.9%;  Pred. No. 3.4e-60;
  Matches  129;  Conservative   26;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :|:|:|| |:|:|||| :: |||||||:  :|||::::  || | | |:|: |||| ||
Db       1030 LAPITAYTQQTRGLLGAIVVSLTGRDKNEQAGQVQVLSSVTQTFLGTSISGVLWTVYHGA 1089

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              | :|:: ||||| |||| : |||||:|| |::|| |:|:|: || |||| |||| :|:
Db       1090 GNKTLAGPKGPVTQMYTSAEGDLVGWPSPPGTKSLDPCTCGAVDLYLVTRNADVIPVRRK 1149

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
               | ||:|||||:|| | ||||| :|| .:||||:|:|||  |||||::|||| ||:
Db       1150 DDRRGALLSPRPLSILKGSSGGPVLCPRGHAVGLFRAAVCARGVAKSIDFIPVESLDVAT 1209

Qy        181 RS 182
              |:
Db       1210 RT 1211
```

```
RESULT 8
POLG_HCVJ6
ID   POLG_HCVJ6     STANDARD;      PRT;  3033 AA.
AC   P26660;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate HC-J6) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11113;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92044440; PubMed=1658196;
RA   Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Iizuka H.,
RA   Machida A., Miyakawa Y., Mayumi M.;
RT   "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
RT   from a human carrier: comparison with reported isolates for conserved
RT   and divergent regions.";
RL   J. Gen. Virol. 72:2697-2704(1991).
CC   -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC       HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC       NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate + N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D00944; BAA00792.1; -.
DR   PIR; JQ1303; JQ1303.
DR   HSSP; P27958; 1HEI.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
```

```
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    733       NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT   CHAIN       734   1010       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1011   1619       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1620   1866       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1867   2017       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2018   3033       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1087   1087       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1111   1111       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1169   1169       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1234   1241       ATP (POTENTIAL).
FT   SITE       1320   1323       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    477    477       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    534    534       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    542    542       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    558    558       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    627    627       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    649    649       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1091   1091       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2038   2038       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2811   2811       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3033 AA;  329165 MW;  F957F5C1A273BE9E CRC64;

  Query Match             74.7%;  Score 712;  DB 1;  Length 3033;
  Best Local Similarity   69.8%;  Pred. No. 5.3e-60;
  Matches  127;  Conservative   29;  Mismatches   26;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :||:|||||r|:|||| :: |:|| || :  '|:|:: |  |:|| |  :|| || |||
Db       1030 LAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTTISGVLWIVYHGA 1089

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              | :|:|  :||| |||:: : ||:|||:| |::|| ' |||:  |||||:||..; ||'
Db       1090 GNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRNADVIPARRR 1149

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
               ,| ||:|:||||:| |||| |||:||| ||| |:|||'||:|.'||::|.||| :|:
Db       1150 GDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFIPVETLDIVT 1209

Qy        181 RS 182
              |:
Db       1210 RS 1211


RESULT 9
PAAD_PSEAE
ID   PAAD_PSEAE     STANDARD;      PRT;   209 AA.
AC   Q9HX08;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable aromatic acid decarboxylase (EC 4.1.1.-).
GN   PA4019.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- SIMILARITY: BELONGS TO THE POLYPRENYL P-HYDROXYBENZOATE /
CC       PHENYLACRYLIC ACID DECARBOXYLASES FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE004818; AAG07406.1; -.
DR   PIR; H83144; H83144.
DR   InterPro; IPR003382; Flavoprotein.
DR   Pfam; PF02441; Flavoprotein; 1.
KW   Hypothetical protein; Lyase; Decarboxylase; Complete proteome.
SQ   SEQUENCE   209 AA;  22367 MW;  01FD081CC495D3F6 CRC64;

  Query Match             9.1%;  Score 87;  DB 1;  Length 209;
  Best Local Similarity   26.2%;  Pred. No. 0.28;
  Matches   55;  Conservative   18;  Mismatches   61;  Indels   76;  Gaps   12;

Qy          8 AQQTRGLLGCIITSLIGRDKNQVEGEVQ-IVSTAAQTFLATCINGVCWTVYHGAGTRTIA 66
              ||   || |::          |.| ||  ::| ||| .:||                 :|
Db         17 AQYGLRLLDCLV---------QEEREVHFLISKAAQLVMAT--------------ETDVA 53

Qy         67 SPKGP----------------VIQMYTNVDKDLVGWPAPQGSRSLTP-----CTCGSSDL 105
               | |                 |:::   |    | ||  | |  |     | | : :
Db         54 LPAKPQAMQAFLTEYCGAAAGQIRVFGQND-----WMAPPASGSSAPNAMVICPCSTGTL 108

Qy        106 ---------YLVTRHADVIPVRRRGDSRGSLLSPR--PIS------YLKGSSGGPLLCPA 148
                      |: | :||    |    |  :| |:  | |       || |: | :: ||
Db        109 SAVATGACNNLIERAADVALKER----RPLVLVPREAPFSSIHLENMLKLSNLGAVILPA 164

Qy        149 GHAVGIFRAAVCTRGVAKAVDFIPVESLET 178
                 | | :      : ,   |:|:    | |
Db        165 --APGFYHQ---PQSVEDLVDFVVARILNT 189


RESULT 10
HHOA_ARATH
ID   HHOA_ARATH     STANDARD;      PRT;   321 AA.
AC   Q9SEL7; O49507;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Protease HhoA, chloroplast precursor (EC 3.4.21.-).
GN   HHOA OR AT4G18370 OR F28J12.30.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
```

```
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lensch M.H.A., Sokolenko A., Herrmann R.G.;
RT   "Identification and characterization of the chloroplast HhoA protease,
RT   a homolog to the bacterial periplasmic protease HhoA.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083488; PubMed=10617198;
RA   Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA   Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA   Harris B., Ansorge W., Brandt P., Grivell L., Rieger M.,
RA   Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA   Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA   Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA   Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA   Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA   Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA   Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA   Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA   Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA   Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA   Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA   De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA   Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA   Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA   Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA   Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,
RA   Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA   Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA   Massenet O., Quigley F., Clabauld G., Muendlein A., Felber R.,
RA   Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA   Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA   Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A.,
RA   Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA   Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Bielke C.,
RA   Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA   Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA   Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA   Sehkon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA   Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA   Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA   Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA   Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA   Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA   Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
RA   Antonoiu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA   Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA   Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA   Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA   Chen E., Marra M., Martienssen R., McCombie W.R.;
RT   "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:769-777(1999).
RN   [3]
RP   SEQUENCE OF 72-82; 96-110; 150-159; 178-211 AND 306-320.
RA   Schubert M., Peterson U., Funk C., Haas B., Schroeder W.P.,
RA   Kieselbach T.;
RT   "The chloroplast lumen from Arabidopsis thaliana.";
RL   Submitted (JUL-2001) to the SWISS-PROT data bank.
CC   -!- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC   -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous
CC       gene model prediction. AT4G18370 and AT4G18375 were originally
CC       fused into a single gene.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF114386; AAF24060.1; -.
DR   EMBL; AL021710; CAA16717.1; ALT_SEQ.
DR   EMBL; AL161548; CAB78839.1; ALT_SEQ.
DR   MEROPS; S01.279; -.
DR   InterPro; IPR001940; Protease2C.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00834; PROTEASES2C.
KW   Hydrolase; Serine protease; Chloroplast; Thylakoid; Transit peptide.
FT   TRANSIT       1     26       CHLOROPLAST (POTENTIAL).
FT   TRANSIT      27     71       THYLAKOID.
FT   CHAIN        72    321       PROTEASE HHOA.
FT   DOMAIN       77     87       POLY-GLU.
FT   ACT_SITE    145    145       CHARGE RELAY SYSTEM (POTENTIAL).
FT   ACT_SITE    186    186       CHARGE RELAY SYSTEM (POTENTIAL).
FT   ACT_SITE    264    264       CHARGE RELAY SYSTEM (POTENTIAL).
FT   CONFLICT     40     40       R -> G (IN REF. 1).
SQ   SEQUENCE   321 AA;  34691 MW;  66DB81E0BD27A7A7 CRC64;

  Query Match             8.8%;  Score 84;  DB 1;  Length 321;
  Best Local Similarity   22.2%;  Pred. No. 0.88;
  Matches   48;  Conservative   26;  Mismatches   60;  Indels   82;  Gaps   11;

Qy         22 LTGRDKNQVEGEVQIVSTAAQTFLATCINGVCW-------TVYH---------------- 58
              ||  :  ::|!                    :|  |       | |,
Db        117 LTDEENGKIEGTG--------------SGFVWDKLGHIVTNYHVIAKLATDQFGLQRCK 161

Qy         59 -----GAGTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHAD 113
                   |||   | :: :: :  : | ||         | | |   |:|:   | :
Db        162 VSLVDAKGTR--FSKEGKIVGL--DPDNDLAVLKIETEGRELNPVVLGTSNDLRVGQSCF 217

Qy        114 VIPVRRRGDSRG--------------SLLSPRPISYLK----------GSSGGPLLCPA 148
               |     |: |                 : ||   |  :           |:||||||
Db        218 AI-----GNPYGYENTLTIGVVSGLGREIPSPNGKSISEAIQTDADINSGNSGGPLLDSY 272

Qy        149 GHAVGIFRAAVCIR---GVAKAVDF-IPVESLETTM 180
              || :|:   |  ||   |::  |:| ||:::  |:
Db        273 GHTIGV-NTAFTRKGSGMSSGVNFAIPIDTVVRTV 307


RESULT 11
AAMP_HUMAN
ID   AAMP_HUMAN     STANDARD;      PRT;   452 AA.
AC   Q13685;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Angio-associated migratory cell protein.
GN   AAMP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=95262124; PubMed=7743515;
RA   Beckner M.E., Krutzsch H.C., Stracke M.L., Williams S.T.,
RA   Gallardo J.A., Liotta L.A.;
RT   "Identification of a new immunoglobulin superfamily protein expressed
RT   in blood vessels with a heparin-binding consensus sequence.";
RL   Cancer Res. 55:2140-2149(1995).
CC   -!- FUNCTION: MAY HAVE A FUNCTION IN MIGRATING CELLS.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN BLOOD VESSELS. STRONGLY EXPRESSED
CC       IN ENDOTHELIAL CELLS, CYTOTROPHOBLASTS, AND POORLY DIFFERENTIATED
CC       COLON ADENOCARCINOMA CELLS FOUND IN LYMPHATICS.
CC   -!- SIMILARITY: Contains 8 WD repeats.
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M95627; AAA69889.1; -.
DR   PIR; I39383; I39383.
DR   Genew; HGNC:18; AAMP.
DR   MIM; 603488; -.
DR   GO; GO:0008201; F:heparin binding activity; TAS.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00400; WD40; 8.
DR   SMART; SM00320; WD40; 8.
DR   PROSITE; PS00678; WD_REPEATS_1; 1.
DR   PROSITE; PS50082; WD_REPEATS_2; 6.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW   Repeat; WD repeat.
FT   DOMAIN       14     18       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN       71     77       POLY-GLU.
FT   REPEAT      107    138       WD 1.
FT   REPEAT      150    180       WD 2.
FT   REPEAT      190    220       WD 3.
FT   REPEAT      231    261       WD 4.
FT   REPEAT      276    306       WD 5.
FT   REPEAT      333    363       WD 6.
FT   REPEAT      374    404       WD 7.
FT   REPEAT      416    446       WD 8.
SQ   SEQUENCE   452 AA;  49015 MW;  DA1413D25EB236C0 CRC64;

  Query Match             8.6%;  Score 82;  DB 1;  Length 452;
  Best Local Similarity   25.3%;  Pred. No. 2;
  Matches   42;  Conservative   13;  Mismatches   47;  Indels   64;  Gaps    9;

Qy         54 WTVYHGAGTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSL-----TPCTCGSSDLYLV 108
                |  ::         |: | :   |  | :   :  | |          | : |
Db        197 WMEWH---------PRAPVLLAGT-ADGNTWMWKVPNGDCKTFQGPNCPATCGR------ 240

Qy        109 TRHADVIPVRRR---GDSRGS-----LLSPRPISYLKGSSG--GPLLCPA---------- 148
                 !:|  :|   |   |:     :    ; || : |  ::  | |
Db        241 -----VLPDGKRAVVGYEDGTIR1WDLKQGSPIHVLKGTEGHQGPLTCVAANQDGSLILT 295

Qy        149 --------------GHAVGIFR----AAVCTRGVAKAVDFIPVESL 176
                           | ||:||    :  : |  :   :   ||,
Db        296 GSVDCQAKLVSAIGKVVGVFRPETVASQPSLGEGEESESNSVESL 341


RESULT 12
Y136_TREPA
ID   Y136_TREPA     STANDARD;      PRT;   485 AA.
AC   O83172;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical lipoprotein TP0136 precursor.
GN   TP0136.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Nichols;
RX   MEDLINE=98332770; PubMed=9665876;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White C., Sutton G.G.,
RA   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA   Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA   Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC       (Potential).
CC   -!- SIMILARITY: BELONGS TO THE TP013X FAMILY OF LIPOPROTEINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE001199; AAC65137.1; ALT_INIT.
DR   TIGR; TP0136; -.
KW   Hypothetical protein; Lipoprotein; Membrane; Signal;
KW   Complete proteome.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    485       HYPOTHETICAL LIPOPROTEIN TP0136.
FT   LIPID        24     24       N-ACYL DIGLYCERIDE (POTENTIAL).
FT   DOMAIN      164    178       GLY/SER-RICH.
FT   DOMAIN      196    210       GLY/SER-RICH.
FT   DOMAIN      253    267       GLY/SER-RICH.
FT   DOMAIN      318    327       POLY-SER.
FT   DOMAIN      444    447       POLY-SER.
SQ   SEQUENCE   485 AA;  48984 MW;  C7A4CEEDC7DC5CED CRC64;

  Query Match             8.6%;  Score 82;  DB 1;  Length 485;
  Best Local Similarity   24.2%;  Pred. No. 2.2;
  Matches   44;  Conservative   13;  Mismatches   65;  Indels   60;  Gaps    8;

Qy         23 TGRDKNQVEGEVQIVSTAAQTFLATCI--NGVCWTVYHGAG---TRTIASPKGPVIQMYT 77
                |  ||      :|  ::|   ||:  |:   |:      :|||   : | |||      :
Db         86 TDSDK-----KVMSIAIDGNTFVLACVPGIGVYKHCVNGAGSSSIGTTASPSTETCSQHA 140

Qy         78 NVDKDLVG------WPAPQGSRSLIPCTC-----GSSDLYLVTRHADVIP--------VR 118
                 |||       |  | |: :   | |     |||       !  ::|
Db        141 T----LVGGTSKPFWLVPGGTGNNGNCGCGGGGGGSSSSSSSCIHIWLVPGGTGNNGNCG 196

Qy        119 RRGDSRGSLLSPRPISYLK-------------------------GSSGGPLLCPAGHA 151
                 |   || |      ::|                        || ||  || |
Db        197 CGGGGGGSSSSSSSCIHIKVENTDEQFLDMGEGYVVTTKHLYTKNGSSSAGPAQCPGGGG 256

Qy        152 VG 153
                |
Db        257 GG 258


RESULT 13
DEG1_ARATH
ID   DEG1_ARATH     STANDARD;      PRT;   437 AA.
AC   O22609; Q9LK85;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Protease Do-like 1, chloroplast precursor (EC 3.4.21.-).
GN   DEGP1 OR DEGP OR AT3G27925 OR K16N12.18.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=98175982; PubMed=9507020;
RA   Itzhaki H., Naveh L., Lindahl M., Cook M., Adam Z.;
RT   "Identification and characterization of DegP, a serine protease
RT   associated with the luminal side of the thylakoid membrane.";
RL   J. Biol. Chem. 273:7094-7098(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
```

```
RX   MEDLINE=20363099; PubMed=10907855;
RA   Kaneko T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
RT   and BAC clones.";
RL   DNA Res. 7:217-221(2000).
RN   [3]
RP   SEQUENCE OF 104-118.
RC   STRAIN=cv. Columbia;
RA   Kieselbach T., Bystedt M., Schroeder W.P.;
RL   Submitted (JUL-2000) to the SWISS-PROT data bank.
CC   -!- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
CC       MAY BE INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IN VIVO,
CC       CAN DEGRADE BETA-CASEIN.
CC   -!- ENZYME REGULATION: INHIBITED BY PHENYLMETHYLSULFONYL FLUORIDE AND
CC       O-PHENANTHROLINE.
CC   -!- SUBCELLULAR LOCATION: BOUND TO LUMINAL SIDE OF THE THYLAKOID
CC       MEMBRANE.
CC   -!- INDUCTION: By heat shock.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC   -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF028842; AAC39436.1; -.
DR   EMBL; AP000371; BAB02539.1; -.
DR   EMBL; AP001302; BAB02539.1; JOINED.
DR   MEROPS; S02.279; -.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001940; Protease2C.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00834; PROTEASES2C.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50106; PDZ; 1.
KW   Hydrolase; Serine protease; Transit peptide; Chloroplast; Thylakoid.
FT   TRANSIT       1      ?       CHLOROPLAST (POTENTIAL).
FT   TRANSIT       ?    103       THYLAKOID.
FT   CHAIN       104    437       PROTEASE DO-LIKE 1.
FT   DOMAIN      152    321       SERINE PROTEASE.
FT   DOMAIN      324    421       PDZ.
FT   ACT_SITE    171    171       CHARGE RELAY SYSTEM (POTENTIAL).
FT   ACT_SITE    201    201       CHARGE RELAY SYSTEM (POTENTIAL).
FT   ACT_SITE    280    280       CHARGE RELAY SYSTEM (POTENTIAL).
FT   CONFLICT     12     23       HSPPSSQLSNST -> SSTLFLHSPPSSHL (IN REF.
FT                                2).
FT   CONFLICT     36     36       V -> I (IN REF. 2).
FT   CONFLICT     54     54       P -> S (IN REF. 2).
FT   CONFLICT     60     60       G -> R (IN REF. 2).
FT   CONFLICT     64     64       G -> D (IN REF. 2).
FT   CONFLICT     68     69       LL -> HF (IN REF. 2).
FT   CONFLICT    355    355       L -> V (IN REF. 2).
FT   CONFLICT    381    381       I -> V (IN REF. 2).
FT   CONFLICT    416    416       Q -> E (IN REF. 2).
SQ   SEQUENCE   437 AA;  46213 MW;  1497B1AB3F5FF2A4 CRC64;

  Query Match             9.4%;  Score 80.5;  DB 1;  Length 437;
  Best Local Similarity   25.6%;  Pred. No. 2.7;
  Matches   44;  Conservative   18;  Mismatches   55;  Indels   55;  Gaps    7;

Qy         56 VYHGAGTRTIASPKGPVIQMY----------------TNVDKDLVGW-----------PA 88
              |  |:|:  :   :| ::  |                |  |  :|||:           :
Db        150 VPQGSGSGFVWDKQGHIVTNYHVIRGASDLRVTLADQTTFDAKVVGFDQDKDVAVLRIDA 209

Qy         89 PQGSRSLTPCTCGSSDLYLV----------------TRHADVIPVRRRGDSRGSLLSPRPI 133
              |:    | |   | |   ||                |    |:   || |  |  : ||
Db        210 PK--NKLRPIPVGVSADLLVGQKVFAIGNPFGLDHTLTTGVISGLRREIS--SAATGRPI 265

Qy        134 SYL--------KGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDF-IPVESL 176
                 :          |:|||||  :|  :.|  |      | :  | | ||:::
Db        266 QDVIQTDAAINPGNSGGPLLDSSGTLIGINTAIYSPSGASSGVGFSIPVDTV 317


RESULT 14
CAC3_BOVIN
ID   CAC3_BOVIN     STANDARD;      PRT;   253 AA.
AC   P05805;
DT   01-NOV-1988 (Rel. 09, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Proproteinase E precursor (Procarboxypeptidase A complex component
DE   III) (Procarboxypeptidase A-S6 subunit III) (PROCPA-S6 III).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 1-25.
RX   MEDLINE=91099520; PubMed=2269366;
RA   Pascual R., Vendrell J., Aviles F.X., Bonicel J., Wicker C.,
RA   Puigserver A.;
RT   "Autolysis of proproteinase E in bovine procarboxypeptidase A ternary
RT   complex gives rise to subunit III.";
RL   FEBS Lett. 277:37-41(1990).
RN   [2]
RP   SEQUENCE OF 14-253, AND DISULFIDE BONDS.
RX   MEDLINE=86220198; PubMed=3519215;
RA   Venot N., Sciaky M., Puigserver A., Desnuelle P., Laurent G.;
RT   "Amino acid sequence and disulfide bridges of subunit III, a
RT   defective endopeptidase present in the bovine pancreatic 6 S
RT   procarboxypeptidase A complex.";
RL   Eur. J. Biochem. 157:91-99(1986).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX   MEDLINE=94222022; PubMed=8168476;
RA   Pignol D., Gaboriaud C., Michon T., Kerfelec B., Chapus C.,
RA   Fontecilla-Camps J.C.;
RT   "Crystal structure of bovine procarboxypeptidase A-S6 subunit III, a
RT   highly structured truncated zymogen E.";
RL   EMBO J. 13:1763-1771(1994).
CC   -!- FUNCTION: DEFECTIVE ELASTASE-LIKE SERINE PROTEASE. DOES NOT SEEM
CC       TO HAVE A PROTEASE ACTIVITY. ITS LIKELY FUNCTION IS TO PROTECT
CC       PROCARBOXYPEPTIDASE A AGAINST DENATURATION IN THE ACIDIC
CC       ENVIRONMENT OF THE RUMINANT DUODENUM.
CC   -!- SUBUNIT: HETEROTRIMER OF SUBUNIT III; CARBOXYPEPTIDASE A AND
CC       CHYMOTRYPSINOGEN C.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Pancreas.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR   PDB; 1FON; 14-OCT-96.
DR   PDB; 1PYT; 27-JAN-97.
DR   MEROPS; S01.983; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Serine protease homolog; Pancreas; Digestion; 3D-structure.
FT   PROPEP        1     11       ACTIVATION PEPTIDE.
FT   CHAIN        12    253       PROPROTEINASE E.
FT   DISULFID     41     57
FT   DISULFID    100    103
FT   DISULFID    140    206
FT   DISULFID    171    187
FT   DISULFID    196    227
```

```
FT   STRAND        23     33
FT   TURN          34     35
FT   STRAND        36     41
FT   STRAND        44     47
FT   TURN          48     49
FT   STRAND        50     53
FT   HELIX         55     57
FT   TURN          60     61
FT   STRAND        64     72
FT   TURN          73     74
FT   STRAND        75     84
FT   TURN          87     88
FT   STRAND        90     92
FT   TURN          94     95
FT   TURN          98     99
FT   HELIX        101    103
FT   STRAND       108    111
FT   TURN         119    120
FT   STRAND       126    126
FT   TURN         130    131
FT   TURN         136    137
FT   STRAND       139    144
FT   TURN         146    147
FT   STRAND       157    157
FT   STRAND       159    166
FT   HELIX        168    171
FT   TURN         172    172
FT   TURN         174    177
FT   HELIX        178    180
FT   TURN         183    184
FT   STRAND       185    188
FT   TURN         196    197
FT   TURN         200    201
FT   STRAND       203    207
FT   TURN         209    210
FT   STRAND       213    221
FT   STRAND       223    223
FT   TURN         224    225
FT   STRAND       226    226
FT   TURN         230    231
FT   STRAND       234    238
FT   HELIX        239    241
FT   HELIX        243    253
SQ   SEQUENCE   253 AA;  27337 MW;  24663724D6AE409C CRC64;

  Query Match             7.9%;  Score 75;  DB 1;  Length 253;
  Best Local Similarity   22.7%;  Pred. No. 4.9;
  Matches   46;  Conservative   23;  Mismatches   72;  Indels   62;  Gaps   10;

Qy         17 CIITSLTGR------DKNQVEGEVQIVS-TAAQTFLATCINGVCWTVYHGAGIRTIA--- 66
              || ||  :  :        |:: :||  |::    |     |:          |      :        ::
Db         57 CISTSRTYQVVLGEYDRSVLEGSEQVIPTNAGDLFVHPLWNSNCVACGNDIALVKLSRSA 116

Qy         67 ------------SPKGPVI------------QMYTNVDKDLVGWPAPQG-SRSLTPCI--- 99
                          | | ::                ::|          | | |      ::| |
Db        117 QLGDKVQLANLPPAGDILPNEAPCYISGWGRLYT-------GGPLPDKLQQALLPVVDYE 169

Qy        100 -CGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPAG-----HAV 152
               |   : : :|     :   :    |::            |    ' ||:| ||:       : |
Db        170 HCSQWDWWGITVKKTM--VCAGGDTR---------SGCNGDSGGPLNCPAADGSWQVHGV 218

Qy        153 GIFRAAVCTRGVAKAVDFIPVES 175
                 | :|     : |   |  | :
Db        219 TSFVSAFGCNTIKKPTVFTRVSA 241


RESULT 15
ZP3_RABIT
ID   ZP3_RABIT      STANDARD;      PRT;   415 AA.
AC   P48833;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Zona pellucida sperm-binding protein 3 precursor (Zona pellucida
DE   glycoprotein ZP3) (Sperm receptor) (Zona pellucida protein C)
DE   (Fragment).
GN   ZP3 OR ZPC.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=95143578; PubMed=7841460;
RA   Harris J.D., Hibler D.W., Fontenot G.K., Hsu K.T., Yurewicz E.C.,
RA   Sacco A.G.;
RT   "Cloning and characterization of zona pellucida genes and cDNAs from
RT   a variety of mammalian species: the ZPA, ZPB and ZPC gene families.";
RL   DNA Seq. 4:361-393(1994).
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U05782; AAA74392.1; -.
DR   PIR; S70401; S70401.
DR   InterPro; IPR001507; Endoglin/CD105.
DR   Pfam; PF00100; zona_pellucida; 1.
DR   SMART; SM00241; ZP; 1.
DR   PROSITE; PS00682; ZP_DOMAIN; 1.
KW   Glycoprotein; Signal; Sulfation; Sperm; Receptor; Transmembrane;
KW   Extracellular matrix; Multigene family.
FT   NON_TER       1      1
FT   SIGNAL       <1     18       POTENTIAL.
FT   CHAIN        19    415       ZONA PELLUCIDA SPERM-BINDING PROTEIN 3.
FT   DOMAIN       19    378       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    379    399       POTENTIAL.
FT   DOMAIN      400    415       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       41    301       ZP.
SQ   SEQUENCE   415 AA;  44987 MW;  77396CF1BAA3F5CB CRC64;

  Query Match             7.8%;  Score 74.5;  DB 1;  Length 415;
  Best Local Similarity   26.7%;  Pred. No. 9.5;
  Matches   31;  Conservative   17;  Mismatches   53;  Indels   15;  Gaps    5;

Qy         55 TVYHGAGTRTIASPKGP-VIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHAD 113
              |||     | | : : |   :     : ::     |    :||   :   | ||: |  |:
Db        271 TVYITCHLRVIPAQQAPDRLNKACSFNQSSSSWAPVEGSADICEC-CGNGDCDLIAGS-- 327

Qy        114 VIPVRRRGDSRGSLLSPRPISYLKGSSGGPL--LCPAGHAVGIFRAAVCTRGVAKA 167
                |: :   :| || |  | ::        : |||  '  ||    |       | ::| |
Db        328 --PMNQNHAARSSLRSRRHVTEEADVTVGPLIFLGKAGDPAG-------TEGLASA 374


Search completed: September 27, 2003, 12:19:26
Job time : 26 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 27, 2003, 11:44:04 ; Search time 42 Seconds
(without alignments)
416.751 Million cell updates/sec

Title: US-09-965-594-1
Perfect score: 953
Sequence: 1 MAPITAYAQQTRGLLGCIIT.....GVAKAVDFIPVESLETTMRS 182

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 99%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	943	99.0	3011	1	GNWVC3	genome polyprotein
2	939	98.5	3011	1	S40770	genome polyprotein
3	927	97.3	3011	1	GNWVCH	genome polyprotein
4	892	93.6	3010	1	GNWVTW	genome polyprotein
5	891	93.5	3010	1	GNWVTC	genome polyprotein
6	887	93.1	3010	1	GNWVCJ	genome polyprotein
7	884	92.8	3010	1	A45573	genome polyprotein
8	864	90.7	3010	1	S18030	genome polyprotein
9	804	84.4	3014	1	JC5620	genome polyprotein
10	714	74.9	3033	1	GNWVJ8	genome polyprotein
11	712	74.7	3033	1	JQ1303	genome polyprotein
12	267.5	28.1	3005	2	T08841	polyprotein - dour
13	255.5	26.8	2970	2	T08839	polyprotein - marm
14	87	9.1	209	2	H83144	probable aromatic
15	82	8.6	452	2	I39383	angio-associated m
16	82	8.6	495	2	B71360	hypothetical prote
17	81.5	8.6	476	2	T48199	heat shock transcr
18	78.5	8.2	620	2	F83976	cytochrome-c oxida
19	78.5	8.2	981	2	T18234	beta transducin ho
20	78	8.2	239	2	H89966	serine proteinase
21	77.5	8.1	398	2	B71284	probable periplasm
22	77	8.1	904	2	AB4212	hypothetical prote
23	76.5	8.0	270	2	T06118	hypothetical prote
24	76.5	8.0	868	2	H81775	aconitate hydratas
25	75	7.9	240	1	CPBOA3	procarboxypeptidas
26	74.5	7.8	415	2	S70401	zona pellucida gly
27	74.5	7.8	755	2	S23441	hypothetical prote
28	74.5	7.8	868	2	C81200	aconitate hydratas
29	74.5	7.8	911	2	JN0821	transferrin-bindin
30	74	7.8	140	2	C72705	hypothetical prote
31	74	7.8	377	2	A75335	hypothetical prote
32	74	7.8	808	2	G86208	protein F22G5.28 [
33	73.5	7.7	356	2	F90978	hypothetical prote
34	73.5	7.7	447	2	S76033	hypothetical prote
35	73.5	7.7	451	2	H82044	C4-dicarboxylate t
36	73.5	7.7	566	2	H84203	phosphate ABC tran
37	73.5	7.7	846	2	I04533	hypothetical prote
38	73.5	7.7	910	2	C81832	transferrin-bindin
39	73.5	7.7	915	2	F81196	transferrin-bindin
40	73.5	7.7	1820	2	A55494	latent transformin
41	73	7.7	239	2	AB9967	serine proteinase
42	73	7.7	354	2	T49806	hypothetical prote
43	73	7.7	433	2	H97199	htrA-like serine p
44	73	7.7	535	2	S65762	chitinase (EC 3.2.
45	72.5	7.6	249	2	A55634	granzyme M (EC 3.4

ALIGNMENTS

RESULT 1
GNWVC3
genome polyprotein - hepatitis C virus (strain HCV-1)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonst
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
C;Accession: A39166; PQ0403; PQ0404
R;Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A;Title: Genetic organization and diversity of the hepatitis C virus.
A;Reference number: A39166; MUID:91172826; PMID:1848704
A;Accession: A39166
A;Molecule type: mRNA
A;Residues: 1-3011 <CHO>
A;Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874
R;Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap
J. Gen. Virol. 73, 1131-1141, 1992
A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship
A;Reference number: PQ0393; MUID:92266871; PMID:1316939
A;Accession: PQ0403
A;Molecule type: genomic RNA
A;Residues: 1577-1633 <CHA>
A;Cross-references: DDBJ:D10128
A;Experimental source: isolates E-b16
A;Accession: PQ0404
A;Status: preliminary
A;Molecule type: genomic RNA
A;Residues: 1577-1633 <CH2>
A;Experimental source: isolates E-b17
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstr
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2

Query Match 99.0%; Score 943; DB 1; Length 3011;
Best Local Similarity 98.4%; Pred. No. 5.7e-81;
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60

```
Db       1026 LAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              |||||||||||||| | ||| : |||.| |||| |||| |||| |||| |||. |||  ||
Db       1086 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              |.||||.|||||'.|||||.|.|||||.|||| .|||| |||||'.|||| .||:|.||
Db       1146 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 1205

Qy        181 RS 182
              ||
Db       1206 RS 1207
```

RESULT 2
S40770
genome polyprotein - hepatitis C virus
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C;Accession: S40770; PC1285
R;Okamoto, H.
submitted to the EMBL Data Library, March 1992
A;Reference number: S40770
A;Accession: S40770
A;Molecule type: genomic RNA
A;Residues: 1-3011 <OKA>
A;Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Isuda,
Jpn. J. Exp. Med. 60, 167-177, 1990
A;Title: The 5'-terminal sequence of the hepatitis C virus genome.
A;Reference number: PC1284; MUID:91013116; PMID:2170712
A;Accession: PC1285
A;Molecule type: genomic RNA
A;Residues: 1-513 <OK2>
A;Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
A;Experimental source: isolate HC-J1
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

  Query Match   98.5%; Score 939; DB 1; Length 3011;
  Best Local Similarity   97.8%; Pred. No. 1.4e-80;
  Matches 178; Conservative   3; Mismatches   1; Indels   0; Gaps   0;

```
Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :|||||||||||||||||'||||'|||||| |||| ||||| |||||| |||| ||||
Db       1026 LAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              |||||||||||||:|||||.:||||.||||.:||||.|||||.|||||||||||.|||'.
Db       1086 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGARSLTPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              :|||||'||||||'|||||  |||||||||  |||||  ||||  |||'||||||  ||||||
Db       1146 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHVVGIFRAAVCTRGVAKAVDFIPVESLETTM 1205

Qy        181 RS 182
              ||
Db       1206 RS 1207
```

RESULT 3
GNWVCH
genome polyprotein - hepatitis C virus (strain H)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstr
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A36814; A41546
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A;Description: Genomic structure of the human prototype strain H of hepatitis C viru
A;Reference number: A36814
A;Accession: A36814
A;Molecule type: genomic RNA
A;Residues: 1-3011 <INC>
A;Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: com
A;Reference number: A41546; MUID:92052256; PMID:1658800
A;Contents: annotation
A;Note: neither amino acid nor nucleotide sequence is given
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruc
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,224

  Query Match   97.3%; Score 927; DB 1; Length 3011;
  Best Local Similarity   96.2%; Pred. No. 1.9e-79;
  Matches 175; Conservative   4; Mismatches   3; Indels   0; Gaps   0;

```
Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :|||||||||||||||||:|||||||||||||||||||||||||||||||| |||||||||||||||||
Db       1026 LAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTATQTFLATCINGVCWTVYHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              |||||||||||||| |||||:|||||||||||||||||||||||||||||||||||||||||
Db       1086 GTRTIASPKGPVIQTYTNVDQDLVGWPAPQGSRSLIPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              |'|||||||||||||||||||||||| |||||:|||||||||||||||||||||:|||||||
Db       1146 GDSRGSLLSPRPISYLKGSSGGPLLCPTGHAVGLFRAAVCTRGVAKAVDFIPVENLETTM 1205

Qy        181 RS 182
              ||
Db       1206 RS 1207
```

RESULT 4
GNWVTW
genome polyprotein - hepatitis C virus (strain Taiwan)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstr
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A40244
R;Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992

```
A;Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
A;Reference number: A40244; MUID:92230206; PMID:1314449
A;Accession: A40244
A;Molecule type: genomic RNA
A;Residues: 1-3010 <CHE>
A;Cross-references: GB:M84754
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

  Query Match             93.6%;  Score 892;  DB 1;  Length 3010;
  Best Local Similarity   90.1%;  Pred. No. 4.1e-76;
  Matches  164;  Conservative   12;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAQTFLATCINGVCWTVYHGA 60
              :|||||:||||| ||||:||||||||||||||| |:|:|  |:|:||||  ||||:||
Db       1026 LAPITAYAQQTRGLFGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVYHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              |::|:| |||: | |||||:||||  ||||:|||||:|||:|||:|||||:||
Db       1086 GSKTLAGPKGPITQMYTNVDQDLVGWHAPQGARSLTPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              |:|||||:|||||:||||||||| :|| ||| ||||| |||:|:|||:|| |:| |
Db       1146 GDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFVPVESMETTM 1205

Qy        181 RS 182
              |'
Db       1206 RS 1207

RESULT 5
GNWVTC
genome polyprotein - hepatitis C virus
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
C;Accession: A38465
R;Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.;
J. Virol. 65, 1105-1113, 1991
A;Title: Structure and organization of the hepatitis C virus genome isolated from human
A;Reference number: A38465; MUID:91140698; PMID:1847440
A;Accession: A38465
A;Molecule type: genomic RNA
A;Residues: 1-3010 <TAK>
A;Cross-references: EMBL:M58335; NID:g329770; PIDN:AAA72945.1; PID:g329771
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,20

  Query Match             93.5%;  Score 891;  DB 1;  Length 3010;
  Best Local Similarity   89.0%;  Pred. No. 5.1e-76;
  Matches  162;  Conservative   15;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :|||||:|||||||||||||||:||||||||||||||:|||| |:||||||:||||||||||
Db       1026 LAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              |::|:|:|:||: '|||||:||||| | |:||||||||||||||||||||||||||
Db       1086 GSKTLAAPKGPITQMYTNVDQDLVGWPKPPGARSLTPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              ||||||||||:||||||||||||| |||||||:|||||||||||||||:||||:||||
Db       1146 GDSRGSLLSPRPVSYLKGSSGGPLLCPFGHAVGIFRAAVCTRGVAKAVDFVPVESMETTM 1205

Qy        181 RS 182
              ||
Db       1206 RS 1207

RESULT 6
GNWVCJ
genome polyprotein - hepatitis C virus (strain J)
N;Contains: capsid protein C; envelope protein M; major envelope protein E; nonstru
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
C;Accession: A39253; PS0086
R;Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; S
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
A;Title: Molecular cloning of the human hepatitis C virus genome from Japanese pat
A;Reference number: A39253; MUID:91088550; PMID:2175903
A;Accession: A39253
A;Molecule type: genomic RNA
A;Residues: 1-3010 <KAT>
A;Cross-references: GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611
R;Kato, N.; Ohkoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65B, 219-223, 1989
A;Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence
A;Reference number: PS0085
A;Accession: PS0086
A;Molecule type: genomic RNA
A;Residues: 2650-2707 <KA2>
A;Experimental source: Japanese isolate
C;Comment: The cleavage sites of this polyprotein have not been determined.
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein;
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,

  Query Match             93.1%;  Score 887;  DB 1;  Length 3010;
  Best Local Similarity   87.9%;  Pred. No. 1.2e-75;
  Matches  160;  Conservative   17;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :|||||:||||||||||||||||||||||:||||::||| |:||||:|||||||||||||
Db       1026 LAPITAYSQQTRGLLGCIITSLTGRDKNQVDGEVQVLSTATQSFLATCVNGVCWTVYHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
```

Db 1085 GSKTLAGPKGPITQMYTNVDQDLVGWPAPPGARSMTPCTCGSSDLYLVTRHADVVPVRRR 1145

Qy 121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180

Db 1146 GDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFIPVESMETTM 1205

Qy 181 RS 182

Db 1206 RS 1207

RESULT 7
A45573
genome polyprotein - hepatitis C virus (strain JT)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C;Accession: A45573
R;Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,
Virus Res. 23, 39-53, 1992
A;Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: s
A;Reference number: A45573; MUID:92295714; PMID:1318627
A;Accession: A45573
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-3010 <TAN>
A;Cross-references: GB:D11168; GB:D01171; NID:g221612; PIDN:BAA01943.1; PID:g221613
A;Experimental source: HCV-JT
A;Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIP:106207)
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 92.8%; Score 884; DB 1; Length 3010;
Best Local Similarity 88.5%; Pred. No. 2.3e-75;
Matches 161; Conservative 15; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60

Db 1026 LAPITAYAQQTRGLLGCIVTSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGA 1085

Qy 61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120

Db 1086 GSKTLAGPKGPITQMYTNVDQDLVGWHAPPGARSLTPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy 121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180

Db 1146 GDGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTM 1205

Qy 181 RS 182

Db 1206 RS 1207

RESULT 8
S18030
genome polyprotein - hepatitis C virus (isolate JK1)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Variety: isolate JK1
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
C;Accession: S18030; S33570; A48332; S18029
R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A;Description: A whole genome of hepatitis C virus cDNA was isolated from a single p
A;Reference number: S18028
A;Accession: S18030
A;Molecule type: genomic RNA
A;Residues: 1-3010 <HON>
A;Cross-references: EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479
A;Experimental source: isolate JK1 from an individual
R;Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.
Arch. Virol. 128, 163-169, 1993
A;Title: Sequence analysis of putative structural regions of hepatitis C virus isol
A;Reference number: A48332; MUID:93119270; PMID:8380522
A;Accession: S33570
A;Molecule type: genomic RNA
A;Residues: 1-547,'T',549-621,'V',623-624,'S',626-652,'DL',655-761,'T',763-782 <HOW
A;Cross-references: EMBL:X61591
A;Note: this sequence is inconsistent with the nucleotide translation
A;Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue
as Trp, and TTC for residue 771 as Ser
A;Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBIP:121748)
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; s
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

Query Match 90.7%; Score 864; DB 1; Length 3010;
Best Local Similarity 86.8%; Pred. No. 1.9e-73;
Matches 158; Conservative 15; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60

Db 1026 LAPITAYSQQTRGLFGCIVTSLTGRDKNQVEGEAQVVSTATQSFLATCVNGVCWTVYHGA 1085

Qy 61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120

Db 1086 GSKTLAGPKGPINQMYTNVDQDLVGWQAPSGAASLTPCTYGSSDLYLVTRHADVIPVRRR 1145

Qy 121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180

Db 1146 GDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTM 1205

Qy 181 RS 182

Db 1206 RS 1207

RESULT 9
JC5620
genome polyprotein - hepatitis C virus (isolate EUH1480)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstr
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C;Accession: JC5620
R;Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997
A;Title: The complete coding sequence of hepatitis C virus genotype 5a, the predomin
A;Reference number: JC5620; MUID:97366593; PMID:9223423

A;Accession: JC5620
A;Molecule type: mRNA
A;Residues: 1-3014 <CHA>
A;Cross-references: GB:Y13184
A;Experimental source: genotype 5a, which predominates in South Africa
A;Note: the translation of the nucleotide sequence is not complete in this paper
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;384-406/Region: hypervariable #status predicted
F;390-730/Product: nonstructural protein NS1 #status predicted <NS1>
F;731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
F;1008-1616/Product: hepacivirin #status predicted <NS3>
F;1231-1236/Region: nucleotide-binding motif A (P-loop)
F;1313-1318/Region: nucleotide-binding motif B
F;1317-1320/Region: DEXH motif
F;1617-1863/Product: nonstructural protein NS4a #status predicted <N4A>
F;1864-2014/Product: nonstructural protein NS4b #status predicted <N4B>
F;2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F;2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 84.4%; Score 804; DB 1; Length 3014;
Best Local Similarity 79.1%; Pred. No. 9.5e-68;
Matches 144; Conservative 22; Mismatches 16; Indels 0; Gaps 0;

Qy 1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
Db 1027 LAPITAYAQQTRGVLGAIVVSLTGRDKNEAEGEVQPLSTATQTFLGTCINGVMWTLFHGA 1086

Qy 61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
Db 1087 GSKTLAGPKGPVVQMYTNVDKDLVGWPSPPGKGSLTPCTCGSADLYLVTRHADVIPARRR 1146

Qy 121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
Db 1147 GDTRASLLSPRPISYLKGSSGGPIMCPSGHVVGVFRAAVCTRGVAKALEFVPVENLETTM 1206

Qy 181 RS 182
Db 1207 RS 1208

RESULT 10
GNWVJ8
genome polyprotein - hepatitis C virus (strain HC-J8)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A40250; PQ0397; PQ0559
R;Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Iizuka, H.; Tanaka, T.; Fukuda, S.;
Virology 188, 331-341, 1992
A;Title: Full-length sequence of a hepatitis C virus genome having poor homology to repc
A;Reference number: A40250; MUID:92230232; PMID:1314459
A;Accession: A40250
A;Molecule type: genomic RNA
A;Residues: 1-3033 <OKA>
A;Cross-references: GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1; PID:g221609
R;Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Poutherer, J.F.; Follett, E.; Yap, P.L
J. Gen. Virol. 73, 1131-1141, 1992
A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A;Reference number: PQ0393; MUID:92268871; PMID:1316939
A;Accession: PQ0397
A;Molecule type: genomic RNA
A;Residues: 2678-2754 <CHA>
A;Cross-references: DDBJ:D10134
A;Experimental source: isolate E-b12
R;Kato, N.; Cotsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno
Biochem. Biophys. Res. Commun. 181, 279-285, 1991
A;Title: Distribution of plural HCV types in Japan.
A;Reference number: PQ0554; MUID:92068204; PMID:1720309

A;Accession: PQ0559
A;Molecule type: mRNA
A;Residues: 2678-2729 <KAT>
A;Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstru
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F;1011-1619/Product: hepacivirin #status predicted <NS3>
F;1234-1241/Region: nucleotide-binding motif A (P-loop)
F;1316-1321/Region: nucleotide-binding motif B
F;1320-1323/Region: DEXH motif
F;1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
F;1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
F;2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2(

Query Match 74.9%; Score 714; DB 1; Length 3033;
Best Local Similarity 70.9%; Pred. No. 3.5e-59;
Matches 129; Conservative 26; Mismatches 27; Indels 0; Gaps 0;

Qy 1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
Db 1030 LAPITAYTQQTRGLLGAIVVSLTGRDKNEQAGQVQVISSVIQTFLGTSISGVLWTVYHGA 1089

Qy 61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
Db 1090 GNKTLAGPKGPVTQMYTSAEGDLVGWPSPPGTKSLDPCTCGAVDLYLVTRNADVIPVRRK 1149

Qy 121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
Db 1150 DDRRGALLSPRPLSTLKGSSGGPVLCSRGHAVGLFRAAVCARGVAKSIDFIPVESLDVAT 1209

Qy 181 RS 182
Db 1210 RT 1211

RESULT 11
JQ1303
genome polyprotein - hepatitis C virus (isolate HC-J6)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nons
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
C;Accession: JQ1303
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyaka
J. Gen. Virol. 72, 2697-2704, 1991
A;Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from
A;Reference number: JQ1303; MUID:92044440; PMID:1658196
A;Accession: JQ1303
A;Molecule type: genomic RNA
A;Residues: 1-3033 <OKA>
A;Cross-references: GB:D00944; NID:g221650; PIDN:BAA00792.1; PID:g221651
A;Experimental source: isolate HC-J6 from a Japanese individual
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase;
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F;1011-1619/Product: hepacivirin #status predicted <NS3>
F;1316-1321/Region: nucleotide-binding motif B
F;1320-1323/Region: DEXH motif
F;1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
F;1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
F;2018-3033/Product: nonstructural protein NS5 #status predicted <NO5>
F;196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,

```
  Query Match             74.7%;  Score 712;  DB 1;  Length 3033;
  Best Local Similarity   69.8%;  Pred. No. 5.4e-59;
  Matches  127;  Conservative   29;  Mismatches   26;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :||||.:||||||||  : |:|:|:| :  :|:|::|    :|| | :|| '.|||:
Db       1030 LAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFIGTIISGVLWTVYHGA 1089

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              | :|:|  :||| ||::: : .| |||:|  ::|'  ||||: ||||:.|:|:'|| |.
Db       1090 GNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRNADVIPARRR 1149

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              || ||:'|||||:| ||||||||:||| ||||:|| |||:. |||::|||:|:|:
Db       1150 GDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFIPVETLDIVT 1209

Qy        181 RS 182
              '|
Db       1210 RS 1211
```

RESULT 12
T08841
polyprotein - douroucouli hepatitis GB virus A
C;Species: douroucouli hepatitis GB virus A
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C;Accession: T08841
R;Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
J. Gen. Virol. 79, 41-45, 1998
A;Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
A;Reference number: Z16486; MUID:98120818; PMID:9460920
A;Accession: T08841
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3005 <ERK>
A;Cross-references: EMBL:AF023425; NID:g2828599; PIDN:AAC40502.1; PID:g2828600
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: polyprotein

```
  Query Match             28.1%;  Score 267.5;  DB 2;  Length 3005;
  Best Local Similarity   33.1%;  Pred. No. 1e-16;
  Matches   60;  Conservative   32;  Mismatches   78;  Indels   11;  Gaps    4;

Qy          2 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
              |':    |:  |     : ||: |:|: : || : :: |:     : | :|||  :  :|::
Db        979 APVVV-MQRGLGFFSVVKTSMLGRDEREHEGSIVVLGTSTTRSMGTCVNGVMYTTFHGSN 1037

Qy         62 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
              ||:| | :||    : :      :  :| : :  '|| ||:   :: :        .
Db       1038 ARTLAGPVGPVNCRWWSPSDDVAVYPLPSGASCLEPCKCGTQSVWCIRN--DGALCHGRL 1095

Qy        122 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGV-------AKAVDFIPVE 174
                 :  |   ||   :|||, |:||   ,| ||  |':      :'  ,||          : : :
Db       1096 SKLVELDLPTHISDFRGSSGSPILCDEGHVVGMM-VSVLHRGVKVTGVRYVKPWHLPKD 1154

Qy        175 S 175
              |
Db       1155 S 1155
```

RESULT 13
T08839
polyprotein - marmoset hepatitis GB virus A
C;Species: marmoset hepatitis GB virus A
C;Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 17-Nov-2000
C;Accession: T08839
R;Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
J. Gen. Virol. 79, 41-45, 1998
A;Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
A;Reference number: Z16486; MUID:98120818; PMID:9460920
A;Accession: T08839
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: genomic RNA
A;Residues: 1-2970 <ERK>
A;Cross-references: EMBL:AF023424; NID:g2828597; PIDN:AAC40501.1; PID:g2828598
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: polyprotein

```
  Query Match             26.8%;  Score 255.5;  DB 2;  Length 2970;
  Best Local Similarity   30.1%;  Pred. No. 1.4e-15;
  Matches   59;  Conservative   36;  Mismatches   68;  Indels   33;  Gaps    6;

Qy          2 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
              |::  :      :|  | : ||:|| |: :   | | :: |:      : |:||| :| |||:
Db        970 APVVVH-HHGKGFFGVVKISMTGWDETEHVGNVVVLGTSTTRSMGTCVNGVMYTTYHGSN 1028

Qy         62 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
              ||:|:  |||    : :      |:  :| | |:: | || |        ::::          |
Db       1029 ARTLAAQMGPVNSRWWSASDDVAVYPLPVGAKCLEPCKCQPQGVWVI-----------RN 1077

Qy        122 DSRGSLLS-----------PRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRG------- 163
              |   |:|                    |  :   :||||  |:||   |||||:    :|  ||
Db       1078 D--GALCHGTLGRTVELDLPAELCDFRGSSGSPILCDEGHAVGML-ISVLHRGSRVTGIR 1134

Qy        164 VAKAVDFIPVESLETT 179
                |  : :| |:: |
Db       1135 YTKPWETLPREAITHT 1150
```

RESULT 14
H83144
probable aromatic acid decarboxylase PA4019 [imported] - Pseudomonas aeruginosa (st
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83144
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic p
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83144
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-209 <STO>
A;Cross-references: GB:AE004818; GB:AE004091; NID:g9950200; PIDN:AAG07406.1; GSPDB:G
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA4019
C;Superfamily: dedF protein

```
  Query Match              9.1%;  Score 87;  DB 2;  Length 209;
  Best Local Similarity   26.2%;  Pred. No. 0.69;
  Matches   55;  Conservative   18;  Mismatches   61;  Indels   76;  Gaps   12;

Qy          8 AQQTRGLLGCIITSLTGRDKNQVEGEVQ-IVSTAAQTFLATCINGVCWTVYHGAGTRTIA 66
              ||    || |::           | | ||  ::| |||  :||                  :|
Db         17 AQYGLRLLDCLV---------QEEREVHFLISKAAQLVMAT-------------ETDVA 53

Qy         67 SPKGP---------------VIQMYTNVDKDLVGWPAPQGSRSLTP-----CTCGSSDL 105
               |  |                ':::   |       | ||  |  |         | |  :  |
Db         54 LPAKPQAMQAFLTEYCGAAAGQIRVFGQND-----WMAPPASGSSAPNAMVICPCSTGIL 108

Qy        106 ---------YLVTRHADVIPVRRRGDSRGSLLSPR--PIS------YLKGSSGGPLLCPA 148
                       |: : |||     |     |   :| ||   : |          || |: | :: ||
Db        109 SAVATGACNNLIERAADVALKER----RPLVLVPREAPFSSIHLENMLKLSNLGAVILPA 164

Qy        149 GHAVGIFRAAVCTRGVAKAVDFIPVESLET 178
                | |      :  |    |||:     | |
Db        165 --APGFYHQ---PQSVEDLVDFVVARILNT 189
```

RESULT 15
I39383
angio-associated migratory cell protein - human
C;Species: Homo sapiens (man)

```
C;Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 21-Jul-2000
C;Accession: I39383
R;Beckner, M.E.; Krutzsch, H.C.; Stracke, M.L.; Williams, S.T.; Gallardo, J.A.; Liotta,
Cancer Res. 55, 2140-2149, 1995
A;Title: Identification of a new immunoglobulin superfamily protein expressed in blood v
A;Reference number: I39383; MUID:95262124; PMID:7743515
A;Accession: I39383
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-452 <RES>
A;Cross-references: GB:M95627; NID:g870802; PIDN:AAA68889.1; PID:g870803
C;Genetics:
A;Gene: GDB:AAMP
A;Cross-references: GDB:4573993
A;Map position: 14q32.1-14q32.1
C;Superfamily: unassigned WD repeat proteins; WD repeat homology
F;148-181/Domain: WD repeat homology <WD1>
F;414-447/Domain: WD repeat homology <WD2>

  Query Match              8.6%;  Score 82;  DB 2;  Length 452;
  Best Local Similarity   25.3%;  Pred. No. 5;
  Matches   42;  Conservative   13;  Mismatches   47;  Indels   44;  Gaps    9;

Qy         54 WTVYHGAGTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSL-----IPCTCGSSDLYLV 108
              :  :|         |: ||:   |  | :   |    |           | |||
Db        197 WMEWH---------PRAPVLLAGT-ADGNTWMWKVPNGDCKTFQGPNCPATCGR------ 240

Qy        109 TRHADVIPVRRR---GDSRGS-----LLSPRPISYLKGSSG--GPLLCPA---------- 148
                  |:|  :|   |  |:     |    ||  |||: |  |||   |
Db        241 -----VLPDGKRAVVGYEDGTIRIWDLKQGSPIHVLKGTEGHQGPLTCVAANQDGSLILI 295

Qy        149 --------------GHAVGIFR----AAVCTRGVAKAVDFIPVESL 176
                           |  :|:||    |:  : |  :  :   : ||
Db        296 GSVDCQAKLVSATTGKVVGVFRPETVASQPSLGEGEESESNSVESL 341


Search completed: September 27, 2003, 12:22:12
Job time : 44 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 27, 2003, 12:28:45 ; Search time 2232 Seconds
(without alignments)
1981.817 Million cell updates/sec

Title: US-09-965-594-1
Perfect score: 953
Sequence: 1 MAPITAYAQQTRGLLGCIIT.....GVAKAVDFIPVESLETTMRS 182

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 99%
Listing first 45 summaries

Command line parameters:
-MODEL=frame-_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09965594/runat_26092003_164343_18769/app_query.fasta_1.327
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=99.9 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594_@CGN_1_1_2810_@runat_26092003_164343_18769 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	106	11.1	984	10	BF304699	BF304699 601888252
	2	103.5	10.9	1199	13	BQ892487	BQ892487 AGENCOURT
	3	101	10.6	515	14	CA023748	CA023748 HZ47E17r
	4	101	10.6	583	12	BM374064	BM374064 EBma03_SQ
c	5	99	10.4	615	12	BJ001625	BJ001625 BJ001625
	6	99	10.4	643	12	BJ024121	BJ024121 BJ024121
	7	99	10.4	754	12	BJ016176	BJ016176 BJ016176
c	8	98.5	10.3	961	10	BF203316	BF203316 601865914
	9	98.5	10.3	1031	14	CB950999	CB950999 AGENCOURT
	10	98.5	10.3	1141	11	AK080545	AK080545 Mus muscu
	11	97.5	10.2	779	10	BF631437	BF631437 HVSMEb001
c	12	96.5	10.1	844	11	CNS09D4S	BX053096 Single re
	13	96	10.1	701	10	BF863244	BF863244 963042C02
c	14	96	10.1	846	10	BF182274	BF182274 601804028
	15	95.5	10.0	901	10	BF307233	BF307233 601891502
c	16	95.5	10.0	958	10	BG420860	BG420860 602452062
	17	95	10.0	407	9	AW785806	AW785806 117260 MA
	18	95	10.0	460	14	CB883286	CB883286 HQ01M02w
	19	94.5	9.9	539	28	BH349665	BH349665 *CH230-65E*
c	20	94	9.9	582	14	CB286751	CB286751 CMD45_C08
c	21	94	9.9	1052	10	BG398041	BG398041 602439571
c	22	94	9.9	1263	13	BQ709745	BQ709745 AGENCOURT
c	23	93.5	9.8	736	12	BI768830	BI768830 603057734
	24	93.5	9.8	938	13	BQ894657	BQ894657 AGENCOURT
	25	93.5	9.8	993	9	AL555424	AL555424 AL555424
	26	93	9.8	457	29	CNS02YQL	AL219990 Tetraodon
	27	93	9.8	470	13	BQ758584	BQ758584 EBma07_SQ
	28	93	9.8	1018	12	BQ054587	BQ054587 AGENCOURT
c	29	93	9.8	1291	10	BE622016	BE622016 6C1440668
	30	92.5	9.7	655	14	CB868789	CB868789 HC09G14w
c	31	92.5	9.7	1008	12	BI755608	BI755608 6C3027112
c	32	92.5	9.7	1411	11	BC020343	BC020343 Homo sapi
c	33	92	9.7	756	14	CD348815	CD348815 UI-M-FY0-
	34	92	9.7	832	10	BG387051	BG387051 602454749
	35	92	9.7	871	10	BG178418	BG178418 602330206
	36	92	9.7	898	29	CNS01VR5	AL169466 Tetraodon
	37	92	9.7	963	10	BF794182	BF794182 602255566
c	38	92	9.7	1329	13	BQ960995	BQ960995 AGENCOURT
c	39	92	9.7	1540	10	BF180599	BF180599 601808704
c	40	91.5	9.6	422	14	CB763743	CB763743 AMGNNUC:S
	41	91.5	9.6	539	10	BE757615	BE757615 212104 MA
	42	91.5	9.6	641	9	AU127824	AU127824 AU127824
c	43	91.5	9.6	691	10	BB632604	BB632604 BB632604
	44	91.5	9.6	701	14	CD262790	CD262790 psMA018xE
c	45	91.5	9.6	844	12	BI198486	BI198486 602760491

ALIGNMENTS

RESULT 1
BF304699/c
LOCUS BF304699 984 bp mRNA linear EST 21-NOV-2000
DEFINITION 601888252F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4122276 5', mRNA sequence.
ACCESSION BF304699
VERSION BF304699.1 GI:11251586
KEYWORDS EST.
SOURCE *Homo sapiens (human)*
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 984)

```
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
 TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL    Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1005  row: g  column: 13
            High quality sequence stop: 646.
FEATURES             Location/Qualifiers
     source          1..984
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4122276"
                     /tissue_type="rhabdomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_17"
                     /note="Organ: muscle; Vector: pCTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      133 a    329 c    351 g    171 t
ORIGIN

Alignment Scores:
Pred. No.:              4.34          Length:        984
Score:                  106.00        Matches:       33
Percent Similarity:     45.24%        Conservative:  5
Best Local Similarity:  39.29%        Mismatches:    24
Query Match:            11.12%        Indels:        22
DB:                     10            Gaps:          5

US-09-965-594-1 (1-182) x BF304699 (1-984)

Qy         86 TrpProAlaProGlnGlySerArgSerLeuThr---ProCysThrCysGlySerSerAsp 104
              :||||:::|||:::   |:|:||:::         ||:|.|||.| |
Db        646 TGGCCCAGTCCACGGCATTCCCGTGCGAGAAGAGGACCGIGTACCTGC------------ 599

Qy        105 LeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArg 124
                         | ||||:!.|||       |:|         ||||: ||||:    |:.:||
Db        598 ------------ACCAGGCACGCAGCGAACATACATGCAAGGAGACGTGGT---TCCCGC 554

Qy        125 GlySerLeuLeuSerProArgPro---------------IleSerTyrLeuLysGlySer 139
              |||   |||||:|          .||||                          :: |||
Db        553 GGGCGCCTCTTGTGGGGAAGACCTCGATGGTGTCCAAGCTGGCGCTGCTGTACTGGAAGT 494

Qy        140 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 159
                        |||:::   :.|                          |||:::
Db        493 CGGCACGCTCCGGTCAGGTGCAGC--------------------TTCCAGCGCCCGGCG 455

Qy        160 CysThrArgGly 163
              :||   :||||:
Db        454 TGCGCCCGAGGA 443

RESULT 2
BQ892487
LOCUS       BQ892487               1199 bp    mRNA    linear   EST 16-AUG-2002
DEFINITION  AGENCOURT_6417538 Lupski_sympathetic_trunk Homo sapiens cDNA clone
            IMAGE:6192708 5', mRNA sequence.
ACCESSION   BQ892487
VERSION     BQ892487.1  GI:22284501
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1199)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James R. Lupski
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13595  row: c  column: 13
            High quality sequence start: 57
            High quality sequence stop: 394.
FEATURES             Location/Qualifiers
     source          1..1199
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6192708"
                     /sex="male"
                     /tissue_type="sympathetic trunk"
                     /dev_stage="adult, 16 yr"
                     /lab_host="DH10B"
                     /clone_lib="Lupski_sympathetic_trunk"
                     /note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
                     NotI; Site_2: SalI; cDNA made by oligo-dT priming.
                     Directionally cloned using the following adaptors:
                     5'-TCGACCCACGCGTCCG-3' and
                     5'-GACTAGTTCTAGATCGCGAGCGGCCGCCCT(15)-3'. Size selected >
                     1 kb for average insert length 1.9 kb. This is a primary
                     library, non-amplified. Library constructed by Life
                     Technologies and donated  by J. Lupski, M.D./Ph.D. (Baylor
                     College of Medicine); available through Life
                     Technologies."
BASE COUNT      255 a    362 c    343 g    211 t     28 others
ORIGIN

Alignment Scores:
Pred. No.:              9.79          Length:        1199
Score:                  103.50        Matches:       41
Percent Similarity:     37.42%        Conservative:  17
Best Local Similarity:  26.45%        Mismatches:    53
Query Match:            10.86%        Indels:        44
DB:                     13            Gaps:          6

US-09-965-594-1 (1-182) x BQ892487 (1-1199)

Qy         54 TrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValIle 73
               |||     :::         |||         |||    :::|||    |||||:::|||
Db        484 TGGGATCCATTTTTAATAAAGGGTGCTCTGTTAATCATGGCGCCCACGGGCCCGCTGATA 543

Qy         74 GlnMetTyrThrAsnValAspLysAspLeuValGlyTrpProAlaProGlnGlySerArg 93
                  :::.||       |||    |||
Db        544 CTTCCATTTACCACATGTGACAGTGACTTT------------------------------ 573

Qy         94 SerLeuThrProCysThr--------------------------CysGlySerSerAsp 104
                        ||||.|                          ||:|||:::   :::
Db        574 ---TGTGCTGCCTGCACAGCACCCCCATGACCGATGTTGGGCTTATGTGGAAACGGCGAG 630

Qy        105 LeuTyr-LeuValThr------------ArgHisAlaAspValIleProValArg----- 118
                ::: |||    |||             :::|||||!|:::|||:::|||
Db        631 GCGTTCATTGGCCACICCCCTCCTATAAAAACACGCCAACGTCGTTCCATGGGGCGGGCT 690

Qy        119 ---------------------------ArgArgGlyAspSerArgGlySerLeuLeu-- 128
                                        |||    :::|||||||:::::|||
Db        691 GGGTGTTTGGCAGGCGCAAGCGGGGTGGGGGCATGGTAGGACTCGGGGGGCGATTCTCTG 750

Qy        129 -------SerProArgProIleSerTyrLeuLys--------------GlySerSerGl 141
```

```
                    |||||||||          ||:                    :::   ||
Db        751 AAAACCCCACCCTCGGCCCACCGATGCGCTAAGCCTCCCTTTACAAGCCACCGCCCGCGG 810

Qy        141 yGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysTh 161
              |  ||||||     |||          |||   |||      ||:      :::
Db        811 CCCCCCCCTAACATCTCCTACCCCIGCGGCGCCGGGGGGGAGACGTGGGCGCATACGGGC 870

Qy        161 rArgGlyValAlaLysAlaValAspPheIleProValGluSer 175
                :::|||||   :||     :::  ::    | ::::
Db        871 TCAGGGCGTTTTAAAGCCCCCGGCCTTCGCGCCGGGCGAAGCA 913

RESULT 3
CA023748
LOCUS       CA023748                515 bp    mRNA    linear   EST 23-OCT-2002
DEFINITION  HZ47E17r HZ Hordeum vulgare subsp. vulgare cDNA clone HZ47E17
            5-PRIME, mRNA sequence.
ACCESSION   CA023748
VERSION     CA023748.1  GI:24301122
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 515)
  AUTHORS   Radchuk,V., Zhang,H., Weschke,W., Potokina,E. and Wobus,U.
  TITLE     Barley ESTs from developing seeds
  JOURNAL   Unpublished
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Corrensstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 515   Std Error: 0.00
            Plate: 47  row: E  column: 17
            Seq primer: M13rev.
FEATURES             Location/Qualifiers
     source          1. .515
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="barke"
                     /db_xref="GABI:277158"
                     /db_xref="taxon:112509"
                     /clone="HZ47E17"
                     /tissue_type="pericarp"
                     /dev_stage="0-7 DAP (days after pollination)"
                     /lab_host="XL10-Gold"
                     /clone_lib="HZ"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); pericarp 0-7
                     DAP(days after pollination). Due to a cloning artefact
                     caused by the kit, in most cases the EcoRI site is NOT
                     present, as well as the EcoRIadapter used for cloning. To
                     excise the insert, restriction sites upstream EcoRI should
                     be used (e.g. BamHI, SalI,PstI). NOTE: Also due to the
                     cloning system used Blue/white selection for recombinats
                     is not 100% reliable.Average insert size is 900 bp"
BASE COUNT       82 a    181 c    160 g     92 t
ORIGIN

Alignment Scores:
Pred. No.:                5.6          Length:          515
Score:                    101.00       Matches:         44
Percent Similarity:       41.96%       Conservative:    16
Best Local Similarity:    30.77%       Mismatches:      53
Query Match:              10.60%       Indels:          31
DB:                       14           Gaps:            6

US-09-965-594-1 (1-182) x CA023748 (1-515)

Qy          8 AlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThrSerLeuThrGlyArgAspLys 27
```

```
              ||||||      |||||||||      !!!!!!      |||   |||   |||
Db         70 GCACAGGCGGCGCGCGGCCTAGCTAGCTGTATTGCACTGTCCACCACCACCAGGCACCAT 129

Qy         28 AsnGlnValGlu-------------------------GlyGluValGlnIleValSer 38
                :::|||:::                         ||:|||:::
Db        130 CATCAAATACC-TCCGGCGAGAGTTTCAGCGATATTTCTGGAGAGCTTGGCGCGGAAGGA 188

Qy         39 ThrAlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrVal------ 56
              ||||||           |||   |||   :::        ||||||
Db        189 GGAGCGGCG----------GCATGGTGTTCTTCTGCTTCCTGGTGGACCAGCGGCGGA 236

Qy         57 TyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyr 76
                  ||||.|   :::|||      |||||   :::|||
Db        237 AGGTGCGGAGCAGCAAGCCGGCGGCGGGCATCTGCTCGCGGTGCGGCGGCTGCGCCAGCG 296

Qy         77 ThrAsnValAspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThr 96
                                       |||      |||   ||||||   :::   :::
Db        297 TGGCGGACA----------------TGGAGACGGCCACCCGGCTCTGCTACCTCCTCA 338

Qy         97 ProCysThrCys------GlySerSerAsp-LeuTyrLeuValThrArgHisAlaAspVa 114
              |||||||||          |||:::       |||:::|||:::   ||||||||   ||
Db        339 CCGTGCACCGCGTCACCTGGCGCGCCATCATCTGCACCTTCTGCGGCGCCATGCTCAACT 398

Qy        114 lIleProValArgArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSe 134
              |:::|||                  :::|||||   |||||   :::::
Db        399 CCTACCG----------------CCACTACCGGCTCTACTAGACCCGCGGTCGTTGAC 440

Qy        134 rTyrLeu 136
              :   :::
Db        441 GCTGATT 447

RESULT 4
BM374064
LOCUS       BM374064                583 bp    mRNA    linear   EST 23-JUL-2002
DEFINITION  EBma03_SQ003_I16_R maternal, 8 DPA, no treatment, cv Optic, EBma03
            Hordeum vulgare subsp. vulgare cDNA clone EBma03_SQ003_I16 5', mRNA
            sequence.
ACCESSION   BM374064
VERSION     BM374064.2  GI:21941288
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 583)
  AUTHORS   Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardie,I.,
            Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
  TITLE     Development of Barley Transcriptome Resources
  JOURNAL   Unpublished
COMMENT     On Jan 10, 2002 this sequence version replaced gi:18117454.
            Contact: Waugh R, Marshall DF
            Genome Dynamics/Computational Biology
            Scottish Crop Research Institute
            Invergowrie, Dundee, DD2 5DA, Scotland, UK
            Tel: 00 44 1382 562731
            Fax: 00 44 1382 562426
            Email: est@scri.sari.ac.uk
            All sequence has a Phred quality score of 20 or over
            Seq primer: M13 reverse.
FEATURES             Location/Qualifiers
     source          1. .583
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Optic"
                     /db_xref="taxon:112509"
                     /clone="EBma03_SQ003_I16"
                     /tissue_type="maternal"
                     /dev_stage="8 DPA"
                     /lab_host="DH10B"
                     /clone_lib="maternal, 8 DPA, no treatment, cv Optic,
                     EBma03"
                     /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
```

```
                Non-normalised library, directionally cloned into pSPORT1.
                Derived from maternal tissue dissected from developing
                grains (8 days post anthesis) in glasshouse grown barley
                plants.  Developed as part of the barley transcriptome
                resources of BBSRC/SEERAD funded cereal IGF (Investigating
                Gene Function) project."
BASE COUNT       92 a    176 c    174 g    141 t
ORIGIN

Alignment Scores:
Pred. No.:              6.6             Length:         583
Score:                  101.00          Matches:        44
Percent Similarity:     41.96%          Conservative:   16
Best Local Similarity:  30.77%          Mismatches:     53
Query Match:            10.60%          Indels:         31
DB:                     12              Gaps:           6

US-09-965-594-1 (1-182) x BM374064 (1-583)

Qy          8 AlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThrSerLeuThrGlyArgAspLys 27
              ||||||      .|||.||||      ||||||      ||   |'|   |'|
Db         11 GCACAGGCGGCGCGCGGCCTAGCTAGCTGTATTGCACTGTCCACCACCACCAGGCACCAT 70

Qy         28 AsnGlnValGlu--------------------------GlyGluValGlnIleValSer 38
              :::|'|:::                          ||||.:::
Db         71 CATCAAATACC-TCCGGCGAGAGTTTCAGCGATATTTCTGGAGAGCTTGGCGCGGAAGGA 129

Qy         39 ThrAlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrVal------ 56
              ||||||      '||   :||   :::       .|||||
Db        130 GGAGCGGCG-----------GCATGGTGTTCTTCTGCTTCCTGGTGGACCAGCGGCGGA 177

Qy         57 TyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyr 76
                 ||||||   :::|||      ;||||;   :::|;|
Db        178 AGGTGCGGAGCAGCAAGCCGGCGGCGGGCATCTGCTCGCGGTGCGGCGGCTGCGCCAGCG 237

Qy         77 ThrAsnValAspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThr 96
                          |||      |.:   ||.|||   :::   :::
Db        238 TGGCGGACA-----------------TGGAGACGGCCACCCGGCTCTGCTACCTCCTCA 279

Qy         97 ProCysThrCys------GlySerSerAsp-LeuTyrLeuValThrArgHisAlaAspVa 114
              ||||||||:        |:|:::      |||:::|||:::   |||:||||   ||
Db        280 CCGTGCACCGCGTCACCTGGCGCGCCATCATCTGCACCTTCTGCGGCGCCATGCTCAAGT 339

Qy        114 lIleProValArgArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSe 133
              |:::|||                        :::||||'   |: |||   :::::
Db        340 CCTACCG-----------------CCACTACCGGCTCTACTAGACCCGCGGTCGTTGAC 391

Qy        134 rTyrLeu 136
              :   :::
Db        392 CCTGATT 398

RESULT 5
BJ001625/c
LOCUS       BJ001625                 615 bp    mRNA    linear   EST 05-DEC-2001
DEFINITION  BJ001625 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA025C02 5',
            mRNA sequence.
ACCESSION   BJ001625
VERSION     BJ001625.1  GI:17364516
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 615)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..615
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-1R"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA025C02"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSA cDNA"
BASE COUNT      140 a    166 c    165 g    144 t
ORIGIN

Alignment Scores:
Pred. No.:              11              Length:         615
Score:                  99.00           Matches:        42
Percent Similarity:     33.77%          Conservative:   9
Best Local Similarity:  27.81%          Mismatches:     50
Query Match:            10.39%          Indels:         50
DB:                     12              Gaps:           7

US-09-965-594-1 (1-182) x BJ001625 (1-615)

Qy         27 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 46
              ||||||   ||||||   :::           :::|||                  |||
Db        511 AAAAATGACGTAGAACCAAAAGACACAGATCCAACCACACATGTTCTGGTTCTACGGGCT 452

Qy         47 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 66
                              ||||||      ||||||   :::      |||      |||
Db        451 -----------------TGTTGGAGAACCTATCACAGTTCCTGCTTTAGAGCAACGGCA 410

Qy         67 SerProLys---------------GlyProValIleGlnMetTyrThrAsnValAspLys 81
              :::   :::               |||||||||:::
Db        409 GCTCTGCGGCGGCGAGGAGCTCCTGGGCCAGTTGTG----------------------- 374

Qy         82 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrPro------------ 97
                                     :::|||   |||      :::|||
Db        373 --------------------ACTCGTGGAGGACGAAGAGCGTCACCCCCGGAGCTGTAGG 335

Qy         98 ---------CysThrCysGlySerSerAspLeuTyrLeuValThrArg------------ 110
                     |||   |||||||||:::         |||||   |||
Db        334 CTGCAGGGATGCGGATGTGGCTCTGCT---------TTGGTTCCTCGTCCTCCTGGATCA 284

Qy        111 --------------------HisAlaAspValIleProValArgArgArgGlyAspSer 123
                                  |||      |||                  ||||||
Db        283 TCTTCCTCACCTGACCTTCCACATCCAGGTGTCGCCAGCGCTGGTCTGACGGGTGATGGG 224

Qy        124 ArgGlySerLeuLeuSerProArg---------ProIleSerTyrLeuLysGlySerSer 140
              ||||||      |||   |||                     |||   |||   |||
Db        223 AGAGGCCGGACAGGCAGCAGTCGGGGGTGAATCTCTGCAGGACGTCTTCACGGCGGATCA 164

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAla 151
              |||||||:|      |||   |||      |||
Db        163 GGAGGACCGACTCGCTGCAGAGCCTCTGCTGCA 131

RESULT 6
BJ024121
LOCUS       BJ024121                 643 bp    mRNA    linear   EST 05-DEC-2001
DEFINITION  BJ024121 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA143D12 3',
            mRNA sequence.
ACCESSION   BJ024121
VERSION     BJ024121.1  GI:17377389
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
```

```
REFERENCE   1  (bases 1 to 643)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..643
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA143D12"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSA cDNA"
BASE COUNT      171 a    148 c    148 g    176 t
ORIGIN

Alignment Scores:
Pred. No.:              11.7          Length:         643
Score:                  99.00         Matches:        42
Percent Similarity:     33.77%        Conservative:   9
Best Local Similarity:  27.81%        Mismatches:     50
Query Match:            10.39%        Indels:         50
DB:                     12            Gaps:           7

US-09-965-594-1 (1-182) x BJ024121 (1-643)

Qy         27 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 46
              |||||  |||||  :::             :::|||
Db        242 AAAAATGACGTAGAACCAAAAGACACAGATCCAACCACACATGTTCTGGTTCTACGGGCT 301

Qy         47 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 66
                                |||||    |||||   :::       ||       |||
Db        302 -----------------TGTTGGAGAACCTATCACAGTTCCTGCTTTAGAGCAACGGCA 343

Qy         67 SerProLys---------------GlyProValIleGlnMetTyrThrAsnValAspLys 81
              :::   :::               ||||||;||:::
Db        344 GCTCTGCGGCGGCGAGGAGCTCCTGGGCCAGTTGTG------------------------ 379

Qy         82 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrPro------------ 97
                                   :::||   |||      :::|||
Db        380 --------------------ACTCGTGGAGGACGAAGAGCGTCACCCCGGAGCTGTAGG 418

Qy         98 ---------CysThrCysGlySerSerAspLeuTyrLeuValThrArg------------ 110
                       |||   ||||||||:::          |||||    |||
Db        419 CTGCAGGGATGCGGATGTGGCTCTGCT---------TTGGTTCCTCGTCCTCCTGGATCA 469

Qy        111 --------------------HisAlaAspValIleProValArgArgArgGlyAspSer 123
                                  |||       |||                 ||||||
Db        470 TCTTCCTCACCTGACCTTCCACATCCAGGTGTCGCCAGCGCTGGTCTGACGGGTGATGGG 529

Qy        124 ArgGlySerLeuLeuSerProArg---------ProIleSerTyrLeuLysGlySerSer 140
              ||||||        |||   |||           |||   |||   |||
Db        530 AGAGGCCGGACAGGCAGCAGTCGGGGGTGAATCTCTGCAGGACGTCTTCACGGCGGATCA 589

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAla 151
              ||||||||     |||   |||       |||
Db        590 GGAGGACCGACTCGCTGCAGAGCCTCTGCTGCA 622

RESULT 7
BJ016176
LOCUS       BJ016176            754 bp    mRNA    linear   EST 05-DEC-2001
DEFINITION  BJ016176 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA025C02 3',
            mRNA sequence.
ACCESSION   BJ016176
VERSION     BJ016176.1  GI:17376695
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 754)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..754
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA025C02"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSA cDNA"
BASE COUNT      194 a    181 c    181 g    198 t
ORIGIN

Alignment Scores:
Pred. No.:              14.4          Length:         754
Score:                  99.00         Matches:        42
Percent Similarity:     33.77%        Conservative:   9
Best Local Similarity:  27.81%        Mismatches:     50
Query Match:            10.39%        Indels:         50
DB:                     12            Gaps:           7

US-09-965-594-1 (1-182) x BJ016176 (1-754)

Qy         27 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 46
              |||||  |||||  :::             :::|||                   |||
Db        242 AAAAATGACGTAGAACCAAAAGACACAGATCCAACCACACATGTTCTGGTTCTACGGGCT 301

Qy         47 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 66
                                |||||    |||||   :::       ||       |||
Db        302 -----------------TGTTGGAGAACCTATCACAGTTCCTGCTTTAGAGCAACGGCA 343

Qy         67 SerProLys---------------GlyProValIleGlnMetTyrThrAsnValAspLys 81
              :::   :::               ||||||;||:::
Db        344 GCTCTGCGGCGGCGAGGAGCTCCTGGGCCAGTTGTG------------------------ 379

Qy         82 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrPro------------ 97
                                   :::||   |||      :::|||
Db        380 --------------------ACTCGTGGAGGACGAAGAGCGTCACCCCGGAGCTGTAGG 418

Qy         98 ---------CysThrCysGlySerSerAspLeuTyrLeuValThrArg------------ 110
                       |||   ||||||||:::          |||||    |||
Db        419 CTGCAGGGATGCGGATGTGGCTCTGCT---------TTGGTTCCTCGTCCTCCTGGATCA 469

Qy        111 --------------------HisAlaAspValIleProValArgArgArgGlyAspSer 123
                                  |||       |||                 ||||||
Db        470 TCTTCCTCACCTGACCTTCCACATCCAGGTGTCGCCAGCGCTGGTCTGACGGGTGATGGG 529

Qy        124 ArgGlySerLeuLeuSerProArg---------ProIleSerTyrLeuLysGlySerSer 140
              ||||||        |||   |||           |||   |||   |||
Db        530 AGAGGCCGGACAGGCAGCAGTCGGGGGTGAATCTCTGCAGGACGTCTTCACGGCGGATCA 589

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAla 151
              ||||||||     |||   |||       |||
```

```
Db        590 GGAGGACCGACTCGCTGCAGAGCCCTCTGCTGCA 622

RESULT 8
BF203316/c
LOCUS       BF203316               961 bp    mRNA    linear   EST 06-NOV-2000
DEFINITION  601865914F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4096578 5',
            mRNA sequence.
ACCESSION   BF203316
VERSION     BF203316.1  GI:11096902
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 961)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM965  row: l  column: 03
            High quality sequence stop: 637.
FEATURES             Location/Qualifiers
     source          1..961
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4096578"
                     /tissue_type="rhabdomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_17"
                     /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      230 a    297 c    300 g    134 t
ORIGIN

Alignment Scores:
Pred. No.:               22.1         Length:          961
Score:                   98.50        Matches:         32
Percent Similarity:      52.44%       Conservative:    11
Best Local Similarity:   39.02%       Mismatches:      32
Query Match:             10.34%       Indels:          7
DB:                      10           Gaps:            2

US-09-965-594-1 (1-182) x BF203316 (1-961)

Qy         87 ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSer-AspLeuTy 106
              |||||||||             ||:::          ::||||:::         ||:::
Db        642 CCGGCTCCTCCTCCTCCTCGGGCA---------TGAGAATGTGGAGCTCAGAGACATGCT 592

Qy        106 rLeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySe 126
                 ::::::        ||||||          |::::|||||||::::     |||
Db        591 GACTCTGAGTGCAGGAGCGGACAGAGATGCGGATCGGAAGCGCGGGGATGCTGATGGGGG 532

Qy        126 rLeuLeuSer-ProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuC 146
                :::||| || ||:||||:||              :::||||    ||||:|
Db        531 CATGCTGTCCTCCTCGTCCCGCGGGGAGTAGGCAAGGGTCTGGAGGAGGAGGG------T 478

Qy        146 ysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaL 166
              |||||||      ||||||           :::|||      |||     |||    ||::
Db        477 GTCCTGCAGTGCTGCTTGTAGGTGCTGTCCCGTCAGCTTTTGTGATGGTGGTGACGCTG 418

Qy        166 ys 166
                ::
Db        417 AA 416

RESULT 9
CB950999
LOCUS       CB950999              1031 bp    mRNA    linear   EST 29-APR-2003
DEFINITION  AGENCOURT_13445496 NIH_MGC_177 Mus musculus cDNA clone
            IMAGE:30316162 5', mRNA sequence.
ACCESSION   CB950999
VERSION     CB950999.1  GI:30205777
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1031)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
             cDNA Library Preparation: Michael Brownstein Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM107  row: b  column: 11
            High quality sequence stop: 333.
FEATURES             Location/Qualifiers
     source          1..1031
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30316162"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /clone_lib="NIH_MGC_177"
                     /note="Organ: liver; Vector: pDNR-LIB; Site_1: SfiI
                     (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); cDNA made
                     by oligo-dT priming and directionally cloned. 5' and 3'
                     adaptors were used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected to contain the 0.5 kb
                     size fraction. Library created in the laboratory of M.
                     Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."
BASE COUNT      235 a    309 c    211 g    275 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:               24.3         Length:          1031
Score:                   98.50        Matches:         47
Percent Similarity:      40.38%       Conservative:    16
Best Local Similarity:   30.13%       Mismatches:      59
Query Match:             10.34%       Indels:          35
DB:                      14           Gaps:            7

US-09-965-594-1 (1-182) x CB950999 (1-1031)

Qy         16 GlyCysIleIleThr--------------------SerLeuThrGlyArgAspLysAsn 28
                ||||:|:::||:                      ||||||    :::    :::
Db        332 GGATGTCTCATTGATTATGTGCCTCAAGTGTGCCACCTGCTCACCCACAAGGTCAGTAT 391

Qy         29 GlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThrCys 48
                 :::::::|||       |||                |||    |||:||     :::
Db        392 GTGATTCAAGGCTATCCCCAAAACAAGAGAGTACATCGGCAGCTTTTCCTTCACTCATTT 451

Qy         49 IleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerPro 68
                :::   |||        |||:::|||    |||                 |||       |||
```

```
Db        452 TTG---GGCACACTGTGGTCCGTGGCACAT----------------ATCGATCGCCCT 490

Qy         69 LysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpProAla 88

Db        491 AAAGGGCCTTTCACAAAAACACTT--------------AACCTCCTTGCCTGGCCTGGC 535

Qy         89 Pro---------------------GlnGlySerArgSerLeuThrProCysThrCysGly 101

Db        536 AT-GTTGGGCAAAAAGAACCGTTTTTGGGTTCGCGCCTCTTGGCCCCCCCCAATTTGGGA 594

Qy        102 SerSerAspLeuTyrLeuValThrArgHisAlaAsp-ValIleProValArgArgGl 121

Db        595 ACCAGTGGC------------ACCACCCATGGGGCTTGTTGTTCCCCGCCTCTCCCGTGG 642

Qy        121 yAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGl 141

Db        643 GCAATTACAAAACNCCCTTAACCGTCCCCTCCCAACAATATTCTTCAAGCGTCCTCTGA 702

Qy        141 y---------GlyProLeuLeuCysProAlaGlyHisAlaValGly 153

Db        703 TTTCCCTAAGTCCCCCCTTTGTTACCCCAGACCACATTTGTGGGA 748

RESULT 10
AK080545
LOCUS       AK080545               1141 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length
            enriched library, clone:A730082L10 product:weakly similar to zinc
            finger protein (fragment) [Mus musculus], full insert sequence.
ACCESSION   AK080545
VERSION     AK080545.1  GI:26348600
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 1141)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .1141
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:A730082L10"
                     /db_xref="taxon:10090"
                     /clone="A730082L10"
                     /tissue_type="cerebellum"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="7 days neonate"
     CDS             <1. .587
                     /note="unnamed protein product; putative
                     weakly similar to zinc finger protein (fragment) [Mus
                     musculus] (PIR|I48722, evidence: FASTY, 50.7%ID,
                     57.6%length, match=601)"
                     /codon_start=3
                     /protein_id="BAC37940.1"
                     /db_xref="GI:26348601"
                     /translation="DSCLPAASPGRSLLTPRGDGFFLKEKLSAARAVGPGPSVAFGVS
                     TRVRGAQAQFGQRRVGACRRSEGLCLSRKPRRRQHVPPVGPHVYGLSGGRRIPPPAGE
                     AQAAGRAPQQVPHPPGRPHGTVVPPQGAAGLLPALAARQVPVGPVRGREGPRRAPRHS
                     PKPVPTALGFSFGQGGPAPPLLAPANGRSVGLAL"
     polyA_signal    1118. .1123
                     /note="putative"
```

```
     polyA_site      1141
                     /note="putative"
BASE COUNT      244 a    316 c    353 g    228 t
ORIGIN

Alignment Scores:
Pred. No.:                27.7          Length:         1141
Score:                    98.50         Matches:        51
Percent Similarity:       34.83%        Conservative:   11
Best Local Similarity:    28.65%        Mismatches:     59
Query Match:              10.34%        Indels:         57
DB:                       11            Gaps:           10

US-09-965-594-1 (1-182) x AK080545 (1-1141)

Qy          13 GlyLeuLeuGlyCysIleIleThrSerLeuThrGlyArgAspLysAsnGlnValGluGly 32
               ||||||   |||                 |||                         ||
Db         267 GGCCTCTCCGGT--------------------GGCAGACGCATCCCGCCGCCGGCGGGA 305

Qy          33 GluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyVal 52
               |||   |||           ||.   ||
Db         306 GAAGCGCAGGCAGCTGGACGCGCGCCGCAG---------------------------- 335

Qy          53 CysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProVal 72
                  ||:   |||      |||                 |||   | :   |.:
Db         336 ------CAAGTGCCGCATCCGCCTGGG--------------CGGCCACATGGAACAGTG 374

Qy          73 IleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpProAlaProGlnGlySer 92
               :::                                          |||.|'|||:::
Db         375 GTG----------------------------------------CCTCCTCAAGGAGCG 392

Qy          93 ArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAla 112
                  |||   |||                    '||      |||
Db         393 GCTGGGCTTCTCCCTGCA----------------------CTCGCAGCTCGCCAAGTT 428

Qy         113 AspValIleProValArg---ArgArgGlyAspSerArgGlySerLeuLeuSerProArg 131
                  |||   ||||||||:   |||   |||      : |           |||||||:::
Db         429 CCTGTTGGACCGGTTAGGGGCCGGGAGGGACCAAGGAGGGCACCGAGACACAGCCCCAAG 488

Qy         132 ProIle-----------------SerTyrLeuLysGlySerSerGlyGlyProLeuLeu 145
                  |||:::                 ,|||:::   :::|||         |||'||||
Db         489 CCGGTTCCTACAGCCTTCGGTTTCTCGTTTGGGCAGGGTGGGCCTGCTCCTCCCCTCCTA 548

Qy         146 CysProAla---GlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyVal 164
                  ||||||   :||   :::||||||:::          ||'|||      '||
Db         549 GCCCCAGCGAATGGGAGGTCTGTAGGGTTG---GCCCTGTAAGTGTGTGAGTGCGGACAC 605

Qy         165 AlaLysAlaVal------AspPheIleProValGluSerLeuGluThrThrMet 180
               :::   |||      :::      ||      || ||'      |.|:::
Db         606 TTGAGGGTGGTTGGCTCTGAGCTGACACCAAGCTCCTCTCTGGCAGGTACACTT 659

RESULT 11
BF631437
LOCUS       BF631437                 779 bp    mRNA    linear   EST 22-OCT-2001
DEFINITION  HVSMEb0015P05f Hordeum vulgare seedling shoot EST library
            HVcDNA0002 (Dehydration stress) Hordeum vulgare subsp. vulgare cDNA
            clone HVSMEb0015P05f, mRNA sequence.
ACCESSION   BF631437
VERSION     BF631437.2  GI:13092107
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 779)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
            ,R.D., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Morex drought-stressed seedling shoot cDNA
            library
  JOURNAL   Unpublished
COMMENT     On Dec 19, 2000 this sequence version replaced gi:11895595.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 455
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 588.
FEATURES             Location/Qualifiers
     source          1..779
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Morex"
                     /db_xref="taxon:112509"
                     /clone="HVSMEb0015P05f"
                     /tissue_type="Seedling shoot"
                     /lab_host="TJC121"
                     /clone_lib="Hordeum vulgare seedling shoot EST library
                     HVcDNA0002 (Dehydration stress)"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1;
                     Seeds were surface sterilized then germinated under axenic
                     conditions in the dark at room temperature on filter paper
                     with water, nystatin and cefotaxime in covered
                     crystallization dishes. Five-day old seedlings were
                     incubated at 90% RH for 24 hr. Shoots were then harvested,
                     total RNA was prepared, poly(A) RNA was purified, one
                     primary unamplified cDNA library was made, 600000 pfu were
                     in vivo excised to give pBluescript SK(-) cDNA phagemids.
                     These steps were performed in the TJ Close laboratory at
                     the University of California, Riverside (Choi, Close,
                     Fenton). Phagemids were plated and picked at the Clemson
                     University Genomics Institute (CUGI) (Begum, Palmer,
                     Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
                     sequencing and sequence analysis were performed at CUGI
                     (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
                     The sequence has been trimmed to remove vector sequence
                     and contains a minimum of 100 bases of phred value 20 or
                     above. For more details on library preparation and
                     sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT      147 a    216 c    311 g    105 t
ORIGIN

Alignment Scores:
Pred. No.:                21            Length:         779
Score:                    97.50         Matches:        47
Percent Similarity:       39.87%        Conservative:   14
Best Local Similarity:    30.72%        Mismatches:     63
Query Match:              10.23%        Indels:         30
DB:                       10            Gaps:           6

US-09-965-594-1 (1-182) x BF631437 (1-779)

Qy          35 GlnIleValSerThrAlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysIrp 54
               ||:::   :::   |||   :::|||           |||           ||'|||
Db          1 CAGCTGCCAACGGCGGCCCCGCGAACAACTACAGGACGGTGC-----------TGCTGG 48

Qy          55 ThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValIle--- 73
               |||   :::   |||   |||   |||   :::      |||      |||
Db         49 ACCCGGTGGAGGGGCGGTGGGCCCCGCTGCCTCTGCTGCCCGGCCCGACCGGGAGCCTGC 108

Qy          74 -----------------GlnMetTyrIhrAsnValAspLysAspLeuValGlyTrpPro 87
                                :::|||                              ||
Db        109 CCCTGTTCTGCCAGGTGGCCGCGGTGGACGGCGCGCAGGGGAGGAAGCGGCTCGTGGTCG 168
```

```
Qy         88 AlaProGlnGlySerArgSer----------LeuThrProCysThrCysGlySerSer 103
Db        169 TCGGCGGGTGGGACCCGGAGTCGTGGGCGCCGACGGACTCGGTGTACGTGTACGACTTCT 228

Qy        104 -AspLeuTyrLeuValThrArg------HisAlaAspValIleProVal------ArgAr 119
Db        229 TGACAGGCGCGTGGCGGAGCGGGGCGCCCATGCCGGGCCCGCGCCGGTCGTTCTTCGCCA 288

Qy        119 gArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySe 139
Db        289 CCGCGGCCGTTGCCGGGGCAGTATACG-TCGCCGGCGGCCACGACGAGGAGAAGAACGCG 347

Qy        139 rSerGly-GlyProLeuLeuCysProAla----------------------GlyHisA 151
Db        348 CTCCGGTCGGCGCTCGCGTACGACCCGGACTCCGACGCGTGGGCCGCGCTCCCGGACATG 407

Qy        151 laValGlyIlePheArgAlaAlaValCysThrArg 162
Db        408 GCGGAGGAGCGCGACGAGCCGCGCGGGCTCTGCGT 442

RESULT 12
CNS09D4S/c
LOCUS       CNS09D4S                 844 bp    mRNA    linear   HTC 08-JAN-2003
DEFINITION  Single read from an extremity of a full-length cDNA clone made from
            Anopheles gambiae total adult females, 3-PRIME end of clone
            FK0AAC30CC12 of strain 6-9 of Anopheles gambiae (African malaria
            mosquito).
ACCESSION   BX053096
VERSION     BX053096.1  GI:27626377
KEYWORDS    HTC.
SOURCE      Anopheles gambiae (African malaria mosquito)
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2003) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Anopheles gambiae"
                     /mol_type="mRNA"
                     /strain="6-9"
                     /db_xref="taxon:7165"
                     /clone="FK0AAC30CC12"
                     /plasmid="pME18S-FL"
                     /note="end : 3-PRIME"
BASE COUNT      197 a    220 c    213 g    214 t
ORIGIN

Alignment Scores:
Pred. No.:                29              Length:          844
Score:                    96.50           Matches:         43
Percent Similarity:       35.87%          Conservative:    23
Best Local Similarity:    23.37%          Mismatches:      53
Query Match:              10.13%          Indels:          55
DB:                       11              Gaps:            10

US-09-965-594-1 (1-182) x CNS09D4S (1-844)

Qy         10 GlnThrArgGlyLeuLeuGlyCysIleIleThrSerLeuThrGlyArgAspLysAsnGln 29
Db        743 CGCACGAAAGCTATAACAAGCTGCGTAATTGCATCCTCCGCCAACCTGGTCTCGCAGAAG 684

Qy         30 ValGluGly------GluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThr 47
Db        683 CTGGGCGGTGCGAAGCAGGTCAACACCGACTCGGTCATGGCGTACGGGCTGTTCGGGCTG 624

Qy         48 CysIleAsnGlyValCysTrpThrValTyrHisGlyAla------GlyThrArgThrIle 65
Db        623 ATCTTTACCGGTCCGTGTCGCACTTCTTCTACAGCTGGCTGGATCGCATCACGAACGATA 564

Qy         66 AlaSerProLys----------------------GlyProValIleGlnMetTyrThr 77
Db        563 CCCGGTTCAAGAAGCTGCTGATGCTGCTGGGCGAGCGGGCCCTCTTTGCGC--------- 513

Qy         78 AsnValAspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrPro 97
Db        512 ----------------------------------------------CGGTCAITACCGCCC 498

Qy         98 CysThrCys----------GlySerSerAspLeuTyrLeuValThrArgHisAlaAsp 113
Db        497 TGTCGCTGTACTTTGTCTCCCGGTTCGAAT--------------ACAAAACACACG--- 456

Qy        114 ValIleProValArgArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIle 133
Db        455 ----------------------ACGAAGCGCTCGGCAATCTGTTCACCCAGTATCGTA 420

Qy        134 SerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGly 153
Db        419 GCATTCTGC---GGGGCAACTGGAAGTGCCTTACCCTGCCCGTATTCATCAACTTCAACT 363

Qy        154 IlePheArgAlaAlaValCysThrArgGlyValAlaLys----------AlaValAsp 169
Db        362 ACATTCCGCCGA------TGTACGCGTGCTGTTTGCCAACATCATCGGATTCTGCTGGAT 309

Qy        170 PheIleProVal 173
Db        308 GGTGTTCCTGTC 297

RESULT 13
BF863244
LOCUS       BF863244                 701 bp    mRNA    linear   EST 19-JAN-2001
DEFINITION  963042C02.xl C. reinhardtii CC-1690, Stress condition I, normalized
            , Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BF863244
VERSION     BF863244.1  GI:12253388
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 701)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Hauser,C.,
            Lefebvre,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 3
  JOURNAL   Unpublished
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .701
                     /organism="Chlamydomonas reinhardtii"
                     /mol_type="mRNA"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Stress condition I,
                     normalized, Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
                     1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr
                     , 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
                     purified from each sample, pooled and cDNA synthesized.
                     The cDNA was directionally cloned into lambda Zap II
                     (Stratagene) in the EcoRI (5') and XhoRI (3') sites.
```

```
                  pBluescript II SK- plasmids were excised from the lambda
                  ZAP clones by superinfection with ExAssist (Stratagene)
                  phage. The library was normalized using method 4 described
                  in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT      173 a    213 c    175 g    140 t
ORIGIN

Alignment Scores:
Pred. No.:               25.4          Length:         701
Score:                   96.00         Matches:        32
Percent Similarity:      40.71%        Conservative:   14
Best Local Similarity:   28.32%        Mismatches:     45
Query Match:             10.07%        Indels:         22
DB:                      10            Gaps:           4

US-09-965-594-1 (1-182) x BF863244 (1-701)

Qy         57 TyrHisGlyAlaGlyThrArgThrIleAlaSerProLys---------GlyProVal 72
Db        171 CACCACCATACCCTGCTCTCACGTTGCTCACACCAAAATTATGCCCATACGGGCCACTA 230

Qy         73 IleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpProAlaProGlnGlySer 92
Db        231 ACAAAGTTACATACACGG------AAGGACCAGCGCGCTTGGCCACCCCCCTTGGAGCCG 284

Qy         93 ArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAla 112
Db        285 AGAAGCCCGACCGCTGGTCCTCTGGGGTCATCCGCATGCTATGCAATCTCCCGCTATCAG 344

Qy        113 AspValIle------------------ProValArgArgArgGlyAspSerArg------ 124
Db        345 GAGATCATTTGCATGTGGCTTTAGTCACCCCAAGAAGAGCCTCGGAGTGGGCATTTATAA 404

Qy        125 -----------------------GlySerLeuLeuSerProArgProIleSerTyrLeu 136
Db        405 GAAGGGGACGGGAATTCCGTTTGCGGAAAAGTGAGCGCCCCAAGGTCTGACCAAGTGCTA 464

Qy        137 LysGlySerSerGlyGlyProLeuLeuCysProAlaGly 149
Db        465 CTCCAAGGCAGCAATGGGAGCCTTTCGCGGTGTGCGGGT 503

RESULT 14
BF182274/c
LOCUS       BF182274               846 bp    mRNA    linear   EST 31-OCT-2000
DEFINITION  601804028F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4035102 5',
            mRNA sequence.
ACCESSION   BF182274
VERSION     BF182274.1  GI:11060416
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 846)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9308  row: g  column: 07
            High quality sequence stop: 696.
FEATURES             Location/Qualifiers
     source          1..846
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4035102"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="7 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam5"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Robin Humphreys,
                     NIH"
BASE COUNT      176 a    218 c    241 g    210 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:               32.5          Length:         846
Score:                   96.00         Matches:        46
Percent Similarity:      47.20%        Conservative:   13
Best Local Similarity:   36.80%        Mismatches:     36
Query Match:             10.07%        Indels:         31
DB:                      10            Gaps:           9

US-09-965-594-1 (1-182) x BF182274 (1-846)

Qy         59 GlyAlaGlyThrArg-ThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAs 78
Db        757 GGTTCTGCTACCAGAAACAGCCTGGATGAAGAAAGGACCA-----------CATCCTTC 710

Qy         78 nValAspLysAspLeuValGlyTrp----------------------ProAlaProGl 90
Db        709 GGTTCTTCAGTCCCAAGTGGGCTGGAGAAAGTGAACACAGAGAAAGAGGACGGTCCCTCA 650

Qy         90 nGlySerArg------SerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuVa 108
Db        649 GTCCCCACGCTTCAGTAGTCTAGACAAGTGTTCTGCTGGA------------------- 610

Qy        108 lThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeuLe 128
Db        609 -ACTAGACACACCT--GTAATCCAGGAGGAACGCNTGGAGGAACAGAGGGACTCC---CT 556

Qy        128 uSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeu---LeuCysPr 147
Db        555 GACCCCACCTCCC---TCCGTCCTAGCGGGCACCTCTCCTGGCCCCACCTCCCTCTGTCC 499

Qy        147 oAlaGlyHis---AlaValGlyIlePheArg-----------AlaAlaValCysThrAr 162
Db        498 TAGTGGGCACCTCTCCCCAGGCACCAGCAGACTGTTACTCCCTTTGGCCCTCTGCACTCT 439

Qy        162 gGlyValAlaLys 166
Db        438 TGGGATGACTGAG 426

RESULT 15
BF307233
LOCUS       BF307233               901 bp    mRNA    linear   EST 21-NOV-2000
DEFINITION  601891502F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4137145 5',
            mRNA sequence.
ACCESSION   BF307233
VERSION     BF307233.1  GI:11254342
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 901)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```

```
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
           Plate: LLCM1044  row: c  column: 02
           High quality sequence start: 6
           High quality sequence stop: 684.
FEATURES             Location/Qualifiers
     source          1. .901
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4137145"
                     /tissue_type="rhabdomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_17"
                     /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      144 a    267 c    329 g    161 t
ORIGIN

Alignment Scores:
Pred. No.:              39.5          Length:         901
Score:                  95.50         Matches:        37
Percent Similarity:     38.46%        Conservative:   8
Best Local Similarity:  31.62%        Mismatches:     26
Query Match:            10.02%        Indels:         45
DB:                     10            Gaps:           5

US-09-965-594-1 (1-182) x BF307233 (1-901)

Qy         52 ValCysTrpThr-ValTyrHis---------------GlyAlaGlyThrArgThrIleAl 66
              :::|||||||::: |||   |||               ||:::||||||
Db        620 ATGTGCTGGAGCGGTCCCGCACGGCCATCCTGAGGCGGGGTCCGGCACACCACAGCGTGG 679

Qy         66 aSerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTr 86
                 |||   |||                         :::   ||||
Db        680 TGGGCAGGACGGTGGAGTGTGTGCA---------------------GTGTCAGGATG 715

Qy         86 pProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTy 106
              |||||||:|   :||
Db        716 GCCCGCCCCATCCGGG---------------------------------------------- 731

Qy        106 rLeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySe 126
                                  :::   :::.||   '||||'      :||||: |
Db        732 ------------------------GTGAGCCTTCGTCTCAGGGGCGTTGGGCGGGGTTC 766

Qy        126 rLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCy 146
              ||||'|    |:|                 :::  ||      :||||. ||    |
Db        767 CCTTTTGGGTCCTGA---------------CGGGCATCTCTCCAGGGGCCGCTGGACTG 810

Qy        146 sProAlaGlyHisAlaValGly------IlePheArgAlaAlaValCys 160
              |||||'|||   |||   |||           ||||'|   :::.||
Db        811 TCCGGCCGGTTCCGCCCGGGGCGGCCACACGGGTCGGGCGGGCCTGTGC 859


Search completed: September 27, 2003, 14:09:39
Job time : 2238 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 27, 2003, 12:22:20 ; Search time 327 Seconds
(without alignments)
1502.440 Million cell updates/sec

Title: US-09-965-594-1
Perfect score: 953
Sequence: 1 MAPITAYAQQTRGLLGCIIT.....GVAKAVEFIPVESLETTMRS 182

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105511

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 99%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09965594/runat_26092003_164342_18746/app_query.fasta_1.327
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=99.9 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594_@CGN_1_1_0_?runat_26092003_164342_18746 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	946	99.3	2058	24	AAD29795	HCV-1 NS3/4a mutan
2	946	99.3	2058	24	ABK75344	Hepatitis C virus
3	946	99.3	2058	25	ABX14410	DNA encoding HCV-1
4	943	99.0	5300	10	AAN92097	Combined open read
5	943	99.0	5360	10	AAN90327	Hepatitis C virus
6	943	99.0	6905	10	AAN92103	Combined open read
7	943	99.0	7310	10	AAN92106	Combined open read
8	943	99.0	7310	10	AAN90336	Composite hepatiti
9	943	99.0	7310	16	AAQ98221	Hepatitis C virus
10	943	99.0	8316	21	AAA75296	cDNA sequence comp
11	943	99.0	9133	20	AAZ07656	Nucleotide sequenc
12	943	99.0	9185	11	AAQ05956	Sense strand of th
13	943	99.0	9185	12	AAQ10566	Hepatitis C virus
14	943	99.0	9185	21	AAA75297	Sense strand of HC
15	943	99.0	9400	13	AAQ21744	Compiled HCV cDNA.
16	943	99.0	9401	17	AAT12710	Hepatitis C virus
17	943	99.0	9401	18	AAT99981	HCV polyprotein co
18	943	99.0	9401	19	AAV09989	HCV polyprotein co
19	943	99.0	9401	24	AAD35043	Hepatitis C virus
20	942	98.8	2061	24	AAD34500	Hepatitis C virus
21	942	98.8	2061	24	AAD31767	Hepatitis C virus
22	940	98.6	9185	20	AAX26737	Nucleotide sequenc
23	940	98.6	9185	20	AAX00459	Hepatitis C virus
24	939	98.5	8316	11	AAQ05955	Hepatitis C virus
25	939	98.5	9502	15	AAQ74770	Hepatitis C virus
26	937	98.3	9646	19	AAV59361	Hepatitis C virus
27	937	98.3	9646	24	ABK87285	cDNA encoding hepa
28	937	98.3	12980	19	AAV59364	Hepatitis C virus
29	937	98.3	12980	24	ABK87286	Hepatitis C virus
30	937	98.3	16622	21	AAZ36212	Nucleotide sequenc
31	936	98.2	630	17	AAT43708	Plasmid pT5His/HIV
32	936	98.2	630	17	AAT42394	HCV insoluble NS3
33	936	98.2	630	18	AAT58401	HCV NS3 protease c
34	936	98.2	810	17	AAT38903	pNB182delta4A HT e
35	936	98.2	810	17	AAT42389	HCV solubilised NS
36	936	98.2	810	18	AAT58396	HCV soluble NS3 pr
37	935	98.1	6299	22	AAF83669	HCV NS34A ORF comp
38	934	98.0	1933	20	AAX23258	HCV NS3 DNA. Hepa
39	934	98.0	8145	20	AAX23259	Plasmid pET-BS(+)/
40	933	97.9	594	21	AAA73335	Hepatitis C virus
41	933	97.9	9365	24	AAD25518	Hepatitis C virus
42	933	97.9	9401	17	AAT41882	Hepatitis C virus
43	933	97.9	9416	19	AAV59378	Hepatitis C virus
44	933	97.9	9416	24	ABK87300	cDNA encoding hepa
45	931	97.7	549	17	AAT43706	Hepatitis C virus

ALIGNMENTS

RESULT 1
AAD29795
ID AAD29795 standard; DNA; 2058 BP.
XX
AC AAD29795;
XX
DT 17-MAY-2002 (first entry)
XX
DE HCV-1 NS3/4a mutant conformational antigen encoding DNA.
XX
KW Hepatitis C virus; NS3/4a antigen; HCV infection; mutant; ds.
XX
OS Hepatitis C virus type 1.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..686

```
FT                   /*tag=  a
FT                   /product= "HCV-1 NS3/4a conformational antigen"
FT                   /note= "CDS does not include stop codon"
FT                   /partial
XX
PN   WO200196875-A2.
XX
PD   20-DEC-2001.
XX
PF   14-JUN-2001; 2001WO-US19369.
XX
PR   15-JUN-2000; 2000US-212082P.
PR   02-APR-2001; 2001US-280811P.
PR   02-APR-2001; 2001US-280867P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Chien DY,  Arcangel P,  Tandeske L,  George-Nasciemento C,  Coit D;
PI   Medina-Selby A;
XX
DR   WPI; 2002-179522/23.
DR   P-PSDB; AAE18689.
XX
PT   Immunoassay solid support useful for detecting hepatitis C virus
PT   infection in a biological sample, comprises at least one of HCV
PT   anti-core antibody and HCV NS3/4a epitope, bound to the support  -
XX
PS   Example 2; Fig 4; 87pp; English.
XX
CC   The present invention relates to hepatitis C virus (HCV) core antigen
CC   and NS (nonstructural) 3/4a antibody combination assay that can detect
CC   both HCV antigens and antibodies present in a sample using a single
CC   solid matrix as well as immunoassay solid supports for use in the assay.
CC   The solid support is useful for detecting HCV infection in a biological
CC   sample. The present sequence is a DNA encoding HCV-1 NS3/4a mutant
CC   conformational antigen. This sequence is used in the exemplification
CC   of the invention.
XX
SQ   Sequence 2058 BP; 419 A; 634 C; 580 G; 425 T; 0 other;

Alignment Scores:
Pred. No.:              1.46e-78        Length:         2058
Score:                  946.00          Matches:        180
Percent Similarity:     100.00%         Conservative:   2
Best Local Similarity:  98.90%          Mismatches:     0
Query Match:            99.27%          Indels:         0
DB:                     24              Gaps:           0

US-09-965-594-1 (1-182) x AAD29795 (1-2058)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 60

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 120

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 180

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 240

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGATCATTGACACCCTGCACTTCC 300

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 360

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 420

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 480

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 540

Qy        181 ArgSer 182
              ||||||
Db        541 AGGTCC 546

RESULT 2
ABK15344
ID   ABK15344 standard; DNA; 2058 BP.
XX
AC   ABK15344;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Hepatitis C virus NS3/4a conformational epitope gene sequence.
XX
KW   Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion;
KW   immunoassay solid support; multiple epitope fusion antigen; MEFA;
KW   non-structural protein; gene; ds.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2058
FT                   /*tag=  a
FT                   /partial
FT                   /product= "HCV NS3/4a conformational epitope"
FT                   /note= "This sequence lacks a stop codon"
XX
PN   WO200196870-A2.
XX
PD   20-DEC-2001.
XX
PF   14-JUN-2001; 2001WO-US19156.
XX
PR   15-JUN-2000; 2000US-212082P.
PR   02-APR-2001; 2001US-280811P.
PR   02-APR-2001; 2001US-280867P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Chien DY,  Arcangel P,  Tandeske L,  George-nasciemento C,  Coit D;
PI   Medina-selby A;
XX
DR   WPI; 2002-090228/12.
DR   P-PSDB; AAU76377.
XX
PT   Immunoassay solid support, useful for detecting hepatitis C virus
PT   infection in biological sample, comprises HCV NS3/4a conformational
PT   epitope and multiple epitope fusion antigen bound to the support  -
XX
PS   Disclosure; Fig 3; 92pp; English.
XX
CC   The present invention relates to a new immunoassay solid support
CC   consisting essentially of at least one hepatitis C virus (HCV) NS3/4a
CC   conformational epitope and a multiple epitope fusion antigen (MEFA),
CC   bound to the support. The NS3/4a conformational epitope and/or
CC   MEFA reacts specifically with anti-HCV antibodies present in a biological
CC   sample from an HCV-infected individual. The immunoassay of the invention
CC   is useful for detecting hepatitis C virus infection in a biological
CC   sample. The method of the invention provides a sensitive, accurate
CC   diagnostic and prognostic tool to provide adequate patient care and to
```

```
CC   prevent transmission of HCV by blood and by blood products, or by
CC   personal contact. Use of NS3/4a conformational epitope in combination
CC   with MEFA, provides a sensitive and reliable method for detecting early
CC   HCV seroconversion. Use of MEFA has the added advantages of decreasing
CC   masking problems, improving sensitivity in detecting antibodies by
CC   allowing a greater number of epitopes on a unit surface area of
CC   substrate, and improving substrate. Detection accuracy is increased and
CC   the incidence of false results is reduced because of the identification
CC   and the use of highly immunogenic HCV antigens which are present during
CC   the early stages of HCV seroconversion. The present nucleic acid sequence
CC   encodes the non-structural protein NS3/4a conformational epitope of the
CC   invention.
XX
SQ   Sequence 2058 BP; 419 A; 633 C; 581 G; 425 T; 0 other;

Alignment Scores:
Pred. No.:               1.46e-78      Length:        2058
Score:                   946.00        Matches:       180
Percent Similarity:      100.00%       Conservative:  2
Best Local Similarity:   98.90%        Mismatches:    0
Query Match:             99.27%        Indels:        0
DB:                      24            Gaps:          0

US-09-965-594-1 (1-182) x ABK15344 (1-2058)

Qy         1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
Db         1 ATGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 60

Qy        21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db        61 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 120

Qy        41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
Db       121 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 180

Qy        61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db       181 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 240

Qy        81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db       241 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGATCATTGACACCCTGCACTTGC 300

Qy       101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
Db       301 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 360

Qy       121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db       361 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 420

Qy       141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db       421 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 480

Qy       161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db       481 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 540

Qy       181 ArgSer 182
Db       541 AGGTCC 546

RESULT 3
ABX14410
ID   ABX14410 standard; DNA; 2058 BP.
XX
AC   ABX14410;
XX
DT   06-MAR-2003  (first entry)
XX
DE   DNA encoding HCV-1 NS3/4a conformational antigen.
XX
KW   Immunoassay solid support; Hepatitis C Virus type-1; HCV-1;
KW   NS3/4a conformational epitope; multiple epitope fusion antigen;
KW   MEFA; anti-HCV antibody; NS3/4a conformational antigen;
KW   HCV infection; mutant; gene; ds.
XX
OS   Hepatitis C virus type 1.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2058
FT                   /*tag= a
FT                   /partial
FT                   /product= "NS3/4a conformational antigen"
FT                   /note= "This sequence lacks a stop codon"
XX
PN   US2002146685-A1.
XX
PD   10-OCT-2002.
XX
PF   14-JUN-2001; 2001US-0881654.
XX
PR   15-JUN-2000; 2000US-212082P.
PR   02-APR-2001; 2001US-280811P.
PR   02-APR-2001; 2001US-280867P.
XX
PA   (CHIE/) CHIEN D Y.
PA   (ARCA/) ARCANGEL P.
PA   (TAND/) TANDESKE L.
PA   (GEOR/) GEORGE-NASCIMENTO C.
PA   (COIT/) COIT D.
PA   (MEDI/) MEDINA-SELBY A.
XX
PI   Chien DY,  Arcangel P,  Tandeske L,  George-Nascimento C,  Coit D;
PI   Medina-Selby A;
XX
DR   WPI; 2003-147573/14.
DR   P-PSDB; ABG72261.
XX
PT   Immunoassay solid support for detecting Hepatitis C Virus infection in
PT   biological samples, comprises Hepatitis C Virus conformational epitope
PT   and multiple epitope fusion antigen -
XX
PS   Disclosure; Fig 3A-3D; 45pp; English.
XX
CC   The present invention relates to immunoassays comprising Hepatitis C
CC   Virus (HCV) NS3/4a conformational epitope and multiple epitope fusion
CC   antigen (MEFA), bound to a solid support. The NS3/4a epitope and/or
CC   the multiple epitope fusion antigen react with anti-HCV antibodies
CC   present in a biological sample from an HCV-infected individual. The
CC   immunoassays and methods of the invention are useful for detecting
CC   HCV infection in a biological sample. The inventive immunoassay solid
CC   support provides a sensitive and reliable method for detecting early
CC   HCV seroconversion. The assays can detect HCV infection caused by any
CC   six known genotypes of HCV. The use of the multiple epitope fusion
CC   proteins decreases masking problems, improves sensitivity in detecting
CC   antibodies by allowing a greater number of epitopes on a unit area
CC   of substrate, and improves selectivity. The present sequence
CC   encodes HCV type 1 (HCV-1) NS3/4a conformational antigen, a mutant of
CC   the HCV-1 NS3/4a polypeptide.
XX
SQ   Sequence 2058 BP; 419 A; 633 C; 581 G; 425 T; 0 other;

Alignment Scores:
Pred. No.:               1.46e-78      Length:        2058
Score:                   946.00        Matches:       180
Percent Similarity:      100.00%       Conservative:  2
Best Local Similarity:   98.90%        Mismatches:    0
Query Match:             99.27%        Indels:        0
DB:                      25            Gaps:          0

US-09-965-594-1 (1-182) x ABX14410 (1-2058)
```

```
Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              ||:|:||'|:||!|:||:|:|||:|||||-::|||:|:|||.:||||:|·|||.|||||
Db          1 ATGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 60

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              ||:|:|||:|||:|:||||:|||:|'|||:|'|||:!'||||'|||! |||:|||
Db         61 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 120

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              :||||||:|||||:|||||||||||:·:||||!.|||·|'||||||||: ||||:||
Db        121 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 180

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              |||:|.||||||:|||||:.||||:||||||:|||:| |||:'||||:.|||:| ||
Db        181 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 240

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              :::|||||||||||:||||||||||||:|||'|.|||·||||||:|||| |||:'||
Db        241 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGATCATTGACACCCTGCACTTGC 300

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              |||||:|!||||:||||||:|'||||::||||.| ||||:! |||||'||| :||||: ||
Db        301 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 360

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              ||:||||:||||:|||||| ||||. |||| ':||| ||:||.:|||| |||.:.||
Db        361 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 420

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              :||||!:||||!·|||:|||||::||||:|||:|||'|||||!|||||!.||||| ||
Db        421 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 480

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              '||||:||||||:|:||||:|:||||·|||'| |||:|·||:::|||||· |||: ||
Db        481 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 540

Qy        181 ArgSer 182
              .|||||
Db        541 AGGTCC 546
```

```
RESULT 4
AAN92097
ID   AAN92097 standard; DNA; 5300 BP.
XX
AC   AAN92097;
XX
DT   25-MAR-2003  (updated)
DT   02-MAR-1990  (first entry)
XX
DE   Combined open reading frames of the hepatitis C virus (HCV) cDNA in
DE   clones 14i 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c,
DE   8f, 33f, 33g and 39c.
XX
KW   Hepatitis C virus; HCV; non-A, non-B hepatitis; NANBH.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..5300
FT                   /*tag=  a
XX
PN   EP318216-A.
XX
PD   31-MAY-1989.
XX
PF   18-NOV-1988;    88EP-0310922.
XX
PR   18-NOV-1987;    87US-0122714.
PR   30-DEC-1987;    87US-0139886.
PR   26-FEB-1988;    88US-0161072.
PR   06-MAY-1988;    88US-0191263.
PR   26-OCT-1988;    88US-0263584.
PR   14-NOV-1988;    88US-0271450.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Houghton M,  Choo QL,  Kuo G;
XX
DR   WPI; 1989-159274/22.
DR   P-PSDB; AAP92041.
XX
PT   Purified hepatitis C virus
PT   - and associated nucleic acids and polypeptide(s)
XX
PS   Claim 3; Figure 26-1, 26-2, 26-3, 26-4, 26-5, 26-6; 139pp; English.
XX
CC   It is a double-stranded nucleotide sequence of the open reading frame
CC   (ORF) (tag a) extending through clones 14i, 11b, 7f, 7e, 8h 33c, 40b,
CC   37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g and 39c of hepatitis C
CC   virus (HCV) cDNA. In creating the composite sequence the following
CC   heterogeneities were considered. Clone 33c contains a sequence
CC   of 800 base pairs which overlaps the cDNAs in clones 40b and 37c. In
CC   clone 33c, as well as in 5 other overlapping clones, nucleotide #789 is
CC   a G. However, in clone 37b the corresponding nucleotide is an A. This
CC   heterogenity may have important ramifications for protein folding.
CC   Nucleotide #2 in clone 8h is a T which may represent a cloning artifact
CC   because the corresponding residue in clone 7e and in 3 other overlapping
CC   clones is an A. Therefore the residue in this position is designated as
CC   an A. The 3'-terminal nucleotide in clone 8f is represented as a T
CC   than a G because the correspondign residue in clone 33f and in 2 other
CC   overlapping clones is a T. The 3' terminal sequence of clone 33f is
CC   represented as ATTC, as is found in the corresponding sequence in clone
CC   33g and in 2 other overlapping clones, rather than as TTGC, as is found
CC   in clone 33f. Residue #4 in clone 33g is designated an A rather than a T
CC   because the corresponding residue in clone 33f and 2 other overlapping
CC   clones is an A. The 3'-terminus of clone 14i is depicted as TA rather
CC   than AA because the corresponding dinucleotide in clone 11b and 3 other
CC   clones is TA. Potential cloning artifacts have been omitted and instead
CC   the corresponding sequences in non-5'-terminal regions of multiple
CC   overlapping clones are shown. AAN92097 could be used as a source of
CC   oligomeric DNA hybridisation probes to detect the presence of HCV
CC   nucleic acids in samples. The polypeptide(s) it encodes could be used as
CC   immuno- assay reagents and vaccines and to generate antibodies useful in
CC   diagnosis and passive immunotherapy for HCV infection/non-A, non-B
CC   hepatitis.
CC   (Updated on 25-MAR-2003 to correct PR field.)
CC   (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ   Sequence 5300 BP; 1047 A; 1606 C; 1515 G; 1130 T; 2 other;

Alignment Scores:
Pred. No.:               8.9e-78        Length:         5300
Score:                   943.00         Matches:        179
Percent Similarity:      100.00%        Conservative:   3
Best Local Similarity:   98.35%         Mismatches:     0
Query Match:             98.95%         Indels:         0
DB:                      10             Gaps:           0

US-09-965-594-1 (1-182) x AAN92097 (1-5300)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              :::|||||||'|||||||||||||||||||||||||||||||||||||||||||||||||
Db        930 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 989

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        990 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1049

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              ||:|||:|||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db       1050 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 1109

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db       1110 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 1169
```

```
Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              :::|||||||||||||||:|||||| |||||||||||:| |||| .||||| ||| :
Db       1170 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 1229

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              |||||||||:|||||| |||||:|||||| |||||:|||||| |||:: |||||.||||
Db       1230 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCCGCGGCGG 1289

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              |||||:|||||:|||||||:||||:|||||:|||||:|||| |:|||||:||| |:|||
Db       1290 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 1349

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              |||||||||||||||||||||||||| |||| |:|||:|||||| ||||| ||
Db       1350 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 1409

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              |||||||:|||||:|||||:|||||:|||||| |||||:|:::|||||.||||: |
Db       1410 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 1469

Qy        181 ArgSer 182
              ||||||
Db       1470 AGGTCC 1475

RESULT 5
AAN90327
ID   AAN90327 standard; cDNA; 5360 BP.
XX
AC   AAN90327;
XX
DT   25-MAR-2003  (updated)
DT   11-NOV-1989  (first entry)
XX
DE   Hepatitis C virus composite probe.
XX
KW   Hepatitis C virus; composite cDNA; probe; vaccine.
XX
OS   Pan troglodytes.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..5360
FT                   /*tag= a
XX
PN   GB2212511-A.
XX
PD   26-JUL-1989.
XX
PF   18-NOV-1988;   88GB-0027024.
XX
PR   18-NOV-1987;   87US-0122714.
PR   30-DEC-1987;   87US-0139686.
PR   26-FEB-1988;   88US-0161072.
PR   26-OCT-1988;   88US-0263584.
XX
PA   (CHIR ) CHIRON CORPORATION.
XX
PT   Houghton M,  Choo QL,  Kuo G;
XX
DR   WPI; 1989-215054/30.
XX
PT   Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT   polypeptide(s) and antibodies for diagnosis, prevention and treatment
PT   of infection.
XX
PS   Disclosure; Fig. 26; 174pp; English.
XX
CC   The sequence shows the composite cDNA sequence derived from the aligned
CC   hepatitis C virus (HCV) cDNA's in clones 14i, 11b, 7f, 7e, 8h, 33c, 40b,
CC   37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g and 39c. The cDNA
CC   encodes antigens which react with antibodies in patients with non-A
CC   non-B hepatitis (NANBH). The cDNA can be used to design probes, or to
CC   synthesise polypeptides, which are used to diagnose HCV-induced NANBH,
CC   to raise antibodies for immunoassay or treatment, or to produce
CC   vaccines. See also AAP90158, AAN90303-26, and AAN90328-36.
CC   (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ   Sequence 5360 BP; 1060 A; 1622 C; 1532 G; 1145 T; 1 other;

Alignment Scores:
Pred. No.:              9.03e-78        Length:         5360
Score:                  943.00          Matches:        179
Percent Similarity:     100.00%         Conservative:   3
Best Local Similarity:  98.35%          Mismatches:     0
Query Match:            98.95%          Indels:         0
DB:                     10              Gaps:           0

US-09-965-594-1 (1-182) x AAN90327 (1-5360)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              :::|||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db        930 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 989

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        990 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1049

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              ||||||||||:||||||||||||||||||||||||||||||:|||||||||||||||||||
Db       1050 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 1109

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1110 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 1169

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              :::|||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db       1170 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 1229

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              |||:||||.|||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1230 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCCGCGGCGG 1289

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              ||||:|||||:|||||||:||||:|||||:|||||:|||||||||||||||||||||:|||
Db       1290 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 1349

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              |||||:|||||:|||||:|||||:|||||||||||||||||||||:|||||||:|||||||
Db       1350 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 1409

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              |||||:|||||:|||||:|||||:|||||| |||| |||||||||:::|||||||||||
Db       1410 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 1469

Qy        181 ArgSer 182
              ||||||
Db       1470 AGGTCC 1475

RESULT 6
AAN92103
ID   AAN92103 standard; DNA; 6905 BP.
XX
AC   AAN92103;
XX
DT   25-MAR-2003  (updated)
DT   02-MAR-1990  (first entry)
XX
DE   Combined open reading frames of the hepatitis C virus (HCV) cDNAs from
DE   clones 12f through 15e.
XX
KW   Hepatitis C virus; HCV; non-A, non-B hepatitis; NANBH.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
```

```
FT   CDS              3..6905
FT                    /*tag= a
XX
PN   EP318216-A.
XX
PD   31-MAY-1989.
XX
PF   18-NOV-1988;   88EP-0310922.
XX
PR   18-NOV-1987;   87US-0122714.
PR   30-DEC-1987;   87US-0139886.
PR   26-FEB-1988;   88US-0161072.
PR   06-MAY-1988;   88US-0191263.
PR   26-OCT-1988;   88US-0263584.
PR   14-NOV-1988;   88US-0271450.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Houghton M,  Choo QL,  Kuo G;
XX
DR   WPI; 1989-159274/22.
DR   P-PSDB; AAP92047.
XX
PT   Purified hepatitis C virus
PT   - and associated nucleic acids and polypeptide(s)
XX
PS   Claim 3; Figure 32-1 - 32-7; 139pp; English.
XX
CC   It is a double-stranded nucleotide sequence of the open reading frame
CC   (ORF) (tag a) extending through clones 12f to 15e of hepatitis C virus
CC   (HCV) cDNA. It can be used to make oligomeric DNA hybridisation probes to
CC   detect the presence of HCV nucleic acids in samples. The polypeptide(s)
CC   it encodes could be used as immunoassay reagents and vaccines and to
CC   generate antibodies useful in diagnosis and passive immunotherapy for
CC   HCV infection/non-A, non-B hepatitis.
CC   (Updated on 25-MAR-2003 to correct PR field.)
CC   (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ   Sequence 6905 BP; 1421 A; 2082 C; 1946 G; 1456 T; 0 other;

Alignment Scores:
Pred. No.:              1.23e-77      Length:          6905
Score:                  943.00        Matches:         179
Percent Similarity:     100.00%       Conservative:    3
Best Local Similarity:  98.35%        Mismatches:      0
Query Match:            98.95%        Indels:          0
DB:                     10            Gaps:            0

US-09-965-594-1 (1-182) x AAN92103 (1-6905)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              :::|||:|:|||||:||||||:||||||||||||:||||  .||||.:||||  :||||: |
Db       1203 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 1262

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              ||:|||||||:||||||:|||||||||||:|:||||:||||  ||||:||||:||||| |
Db       1263 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1322

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              |||:||||:|||||:||||:.||||: ||||:|||||| ||| ||||||:|||:|:|
Db       1323 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 1382

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              ||||:||||:||||||||||||||||||:|||:||||||:|||:|||||:||||||:|||
Db       1383 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 1442

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              :::||:|||||:|||:|||||:||||:||||  |.|||:||||||:.|||:||||||  ||
Db       1443 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 1502

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              |||:|||||:||||:|||||||||||||| ||||| |||||  ||||||||.|||:|||
Db       1503 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 1562

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              |||||||||||||||:||||||||||||||:||||||||||||||:||||| ||||||||
Db       1563 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 1622

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              ||||:||||:||||||||||||||||||||:|||||||||||||||||||||||:||||
Db       1623 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 1682

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              |||:||||| |||||||:||||||||||||||||||:|||||:::.||||||:|||||
Db       1683 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 1742

Qy        181 ArgSer 182
              ||||||
Db       1743 AGGTCC 1748

RESULT 7
AAN92106
ID   AAN92106 standard; DNA; 7310 BP.
XX
AC   AAN92106;
XX
DT   25-MAR-2003  (updated)
DT   02-MAR-1990  (first entry)
XX
DE   Combined open reading frames of the hepatitis C virus (HCV) cDNAs from
DE   clones K9-1 through 15e.
XX
KW   Hepatitis C virus; HCV; non-A, non-B hepatitis; NANBH.
XX
OS   Hepatitis C virus.
XX
FH   Key              Location/Qualifiers
FT   CDS              3..7310
FT                    /*tag= a
XX
PN   EP318216-A.
XX
PD   31-MAY-1989.
XX
PF   18-NOV-1988;   88EP-0310922.
XX
PR   18-NOV-1987;   87US-0122714.
PR   30-DEC-1987;   87US-0139886.
PR   26-FEB-1988;   88US-0161072.
PR   06-MAY-1988;   88US-0191263.
PR   26-OCT-1988;   88US-0263584.
PR   14-NOV-1988;   88US-0271450.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Houghton M,  Choo QL,  Kuo G;
XX
DR   WPI; 1989-159274/22.
DR   P-PSDB; AAP92050.
XX
PT   Purified hepatitis C virus
PT   - and associated nucleic acids and polypeptide(s)
XX
PS   Claim 3; Figure 47-1 - 47-8; 139pp; English.
XX
CC   It is a double-stranded nucleotide sequence of the open reading frame
CC   (ORF) (tag a) extending through clones K9-1 to 15e of hepatitis C virus
CC   (HCV) cDNA. It can be used to make oligomeric DNA hybridisation probes to
CC   detect the presence of HCV nucleic acids in samples. The polypeptide(s)
CC   it encodes could be used as immunoassay reagents and vaccines and to
CC   generate antibodies useful in diagnosis and passive immunotherapy for
CC   HCV infection/non-A, non-B hepatitis.
CC   (Updated on 25-MAR-2003 to correct PR field.)
CC   (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ   Sequence 7310 BP; 1491 A; 2217 C; 2058 G; 1540 T; 4 other;
```

Alignment Scores:

Pred. No.:	1.32e-77	Length:	7310
Score:	943.00	Matches:	179
Percent Similarity:	100.00%	Conservative:	3
Best Local Similarity:	98.35%	Mismatches:	0
Query Match:	98.95%	Indels:	0
DB:	10	Gaps:	0

US-09-965-594-1 (1-182) x AAN92106 (1-7310)

```
Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1728 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 1787

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1788 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1847

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1848 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 1907

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1908 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 1967

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1968 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 2027

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2028 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 2087

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2088 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 2147

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2148 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 2207

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2208 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 2267

Qy        181 ArgSer 182
              ||||||
Db       2268 AGGTCC 2273
```

RESULT 8
AAN90336
ID AAN90336 standard; DNA; 7310 BP.
XX
AC AAN90336;
XX
DT 25-MAR-2003 (updated)
DT 19-JUL-2001 (updated)
DT 01-NOV-1989 (first entry)
XX
DE Composite hepatitis C virus (HCV) cDNA.
XX
KW Hepatitis C virus; cDNA; clone 15e; clone k9-1; probe; vaccine; ds.
XX
OS Pan troglodytes.
XX
PN GB2212511-A.
XX
PD 26-JUL-1989.
XX
PF 18-NOV-1988; 88GB-0027024.
PR 18-NOV-1987; 87US-0122714.
PR 30-DEC-1987; 87US-0139886.
PR 26-FEB-1988; 88US-0161072.
PR 26-OCT-1988; 88US-0263584.
XX
PA (CHIR) CHIRON CORPORATION.
XX
PI Houghton M, Choo QL, Kuo G;
XX
DR WPI; 1989-215054/30.
DR P-PSDB; AAP90288.
XX
PT Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT polypeptide(s) and antibodies for diagnosis, prevention and treatment
PT of infection.
XX
PS Disclosure; fig 47; 235pp; English.
XX
CC The sequence shows a composite hepatitis C virus (HCV) cDNA, derived by
CC aligning clones k9-1 through 15e in 5'-3' direction. The cDNA
CC encodes antigens which react with antibodies in patients with non-A
CC non-B hepatitis (NANBH). The cDNA can be used to design probes, or to
CC synthesise polypeptides, which are used to diagnose HCV-induced NANBH,
CC to raise antibodies for immunoassay or treatment, or to produce
CC vaccines. See also AAP90288, and AAN90303-35.
CC (N.B. This record was resubmitted to correct errors in the sequence.)
CC (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ Sequence 7310 BP; 1495 A; 2216 C; 2058 G; 1539 T; 0 other;

Alignment Scores:

Pred. No.:	1.32e-77	Length:	7310
Score:	943.00	Matches:	179
Percent Similarity:	100.00%	Conservative:	3
Best Local Similarity:	98.35%	Mismatches:	0
Query Match:	98.95%	Indels:	0
DB:	10	Gaps:	0

US-09-965-594-1 (1-182) x AAN90335 (1-7310)

```
Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1728 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 1787

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1788 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1847

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1848 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 1907

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1908 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 1967

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1968 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 2027

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2028 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 2087

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2088 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 2147

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2148 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 2207

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
```

```
              ||||||||.|.||||||||||||||||.|||||.||||' ::::|.||| ::|||.'|
Db       2208 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 2267

Qy        181 ArgSer 182
              ||||||
Db       2268 AGGTCC 2273

RESULT 9
AAQ98221
ID   AAQ98221 standard; cDNA to mRNA; 7310 BP.
XX
AC   AAQ98221;
XX
DT   25-MAR-2003  (updated)
DT   15-AUG-1996  (first entry)
XX
DE   Hepatitis C virus clone genome.
XX
KW   Hepatitis C virus; HCV; antigen; detection; diagnosis; vaccine;
KW   antibodies; immunoprophylaxis; sera; serum; ds.
XX
OS   Hepatitis C virus.
XX
PN   US5443965-A.
XX
PD   22-AUG-1995.
XX
PF   05-APR-1991;   91US-0681703.
XX
PR   06-APR-1990;   90US-0505611.
PR   09-OCT-1990;   90US-0594854.
XX
PA   (GENE-) GENELABS INC.
XX
PI   Kim JP,  Moeckli R,  Reyes GR;
XX
DR   WPI; 1995-302120/39.
XX
PT   New nucleic acids encoding hepatitis C virus antigens - used to
PT   develop prods. for detection of HCV-infected sera and prodn. of
PT   vaccines and anti-HCV antibodies.
XX
PS   Example 4; Figure 11; 71pp; English.
XX
CC   Hepatitis C virus (HCV) antigens can be used for detecting HCV
CC   infected sera and individuals infected with HCV.  They can also be
CC   used in an anti-HCV vaccine or for the production of anti-HCV
CC   antibodies which can be used for passive immunoprophylaxis.  The
CC   antigens consistently identify more HCV positive serum samples with
CC   a high degree of specificity.  See AAQ98202-14 and AAR81939-51.
CC   (Updated on 25-MAR-2003 to correct PF field.)
CC   (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ   Sequence 7310 BP; 1494 A; 2217 C; 2060 G; 1539 T; 0 other;

Alignment Scores:
Pred. No.:              1.32e-77        Length:         7310
Score:                  943.00          Matches:        179
Percent Similarity:     100.00%         Conservative:   3
Best Local Similarity:  98.35%          Mismatches:     0
Query Match:            98.95%          Indels:         0
DB:                     16              Gaps:           0

US-09-965-594-1 (1-182) x AAQ98221 (1-7310)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              :::|||.|||||||||||||||||||| ||||||||||||'||| .||||'|||| |
Db       1728 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTGCTAGGGTGCATAATCACC 1787

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              |||||||||||||||||:|||||.|||||||||| |||||| ||||||||||.||||| |
Db       1788 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1847

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              |||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1848 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 1907

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              |||||||:|||||:||||:|||||||||||||||||||||||||||||||||||| |||||
Db       1908 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 1967

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              :::|||||||||||:|||||:|||||:|||||||||||||||||:||||'||||:|||||||
Db       1968 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACGCTGCACTTGC 2027

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              |||||:|||||| ||||:|||||||||||:|||||||||||||||||||||'||||||||||
Db       2028 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 2087

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              ||||:|||||||||||||:||||||'||||||||||a|||||||||||||||||||||||||
Db       2088 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 2147

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              ||||:|:||||:-||||||||||||||||:|||||.||||||||||||||||||||||||
Db       2148 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 2207

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              ||||:||||'||||||||||||||||:||||,|||||||||:::|||||||||||||||||
Db       2208 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 2267

Qy        181 ArgSer 182
              ||||:
Db       2268 AGGTCC 2273

RESULT 10
AAA75296
ID   AAA75296 standard; cDNA; 8316 BP.
XX
AC   AAA75296;
XX
DT   15-JAN-2001  (first entry)
XX
DE   cDNA sequence compiled Hepatitis C virus cDNA clones.
XX
KW   Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
KW   viral infectivity; viral replication; ds.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..8316
FT                   /*tag= a
FT                   /note= "partial sequence; no termination codon given"
XX
PN   EP1034785-A2.
XX
PD   13-SEP-2000.
XX
PF   16-MAR-1990; 2000EP-0109602.
XX
PR   17-MAR-1989;   89US-0325338.
PR   20-APR-1989;   89US-0341334.
PR   18-MAY-1989;   89US-0355002.
PR   16-MAR-1990;   90EP-0302866.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Houghton M,  Choo Q,  Kuo G;
XX
DR   WPI; 2000-566891/53.
DR   P-PSDB; AAB18540.
XX
PT   Novel composition comprising a hepatitis C virus antisense
PT   polynucleotide which is complementary to or corresponds to a sense
PT   strand of the virus genome, and selectively hybridises to it -
```

```
XX
PS   Example; Fig 16; 75pp; English.
XX
CC   The specification describes a pharmaceutical composition which
CC   comprises a hepatitis C virus (HCV) antisense polynucleotide. The
CC   HCV is characterized by a positive stranded RNA genome which has
CC   40% homology at the polypeptide level to a HCV polyprotein. The
CC   antisense polynucleotide binds to cellular polynucleotides which
CC   enhance and/or are required for viral infectivity, replicative
CC   ability or chronicity. The antisense polynucleotides may also be
CC   designed to bind with high specificity, to be of increased stability,
CC   to be stable and to have low toxicity. The composition also comprises
CC   an agent which causes viral RNA to be inactive. The composition
CC   is used for preventing HCV replication in a system. The present
CC   sequence represents a novel HCV cDNA sequence, which is used in the
CC   course of the invention.
XX
SQ   Sequence 8316 BP; 1671 A; 2529 C; 2345 G; 1771 T; 0 other;

Alignment Scores:
Pred. No.:              1.55e-77     Length:        8316
Score:                  943.00       Matches:       179
Percent Similarity:     100.00%      Conservative:  3
Best Local Similarity:  98.35%       Mismatches:    0
Query Match:            98.95%       Indels:        0
DB:                     21           Gaps:          0

US-09-965-594-1 (1-182) x AAA75296 (1-8316)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              ::: ||||: |||| |.|||||.:|||| :||||: |||| |||| :||||: |||.
Db       2734 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 2793

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              |||:||||||:|||||::||||| |||||:|||| |||  |||| ||||:|||| |||
Db       2794 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 2853

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              ||:|:||||||:|||||:.|||| | ||||:|||| :||||, |||| :|||| :|||
Db       2854 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 2913

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              || : ||||| |:|||| ! ||||: |||| ! ||||: |||| :.||||: |||.: ||
Db       2914 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 2973

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              ::::: |||||::.||||: :||||| |||||:.|||| ! :|||| ||||| :.||| .|
Db       2974 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 3033

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              :|||||::||||:.:|||||:|||:||||.||||:|||:|||||, |||.:||||:|||||:
Db       3034 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 3093

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              |:|||||||||||||||||:||||:||||:||||,:|||||':|||:||| |||||:
Db       3094 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 3153

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              |:|||||:||||'||||||::||||::::||||.:||||' |||||.||| |:||| :
Db       3154 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 3213

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              ||||||||||||||||||:|||||.||||:|||| |||||::::||||.||||||'|||
Db       3214 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 3273

Qy        181 ArgSer 182
              ||||||
Db       3274 AGGTCC 3279

RESULT 11
AAZ07656
ID   AAZ07656 standard; DNA; 9133 BP.
XX
AC   AAZ07656;
XX
DT   20-MAR-2003  (updated)
DT   08-NOV-1999  (first entry)
XX
DE   Nucleotide sequence of HCV-1 ORF.
XX
KW   Hepatitis C virus; HCV; J1; J7; HCV-1; non-A, non-B HCV; NANBH;
KW   HCV infection; vaccine; ds.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             268..9132
FT                   /*tag= b
FT                   /transl_except= (pos:1588..1589; aa:Leu)
FT                   /note= "this codon has an apparent 1 nucleotide deletion,
FT                          which alters the reading frame"
FT                   /transl_except= (pos:1647..1650; aa:Pro)
FT                   /note= "this codon has an apparent 1 nucleotide
FT                          insertion, which alters the reading frame; this
FT                          insertion is not indicated in the sequence
FT                          present in the formal sequence listing of the
FT                          specification"
XX
PN   EP939128-A2.
XX
PD   01-SEP-1999.
XX
PF   17-SEP-1990;   99EP-0101746.
XX
PR   15-SEP-1989;   89US-0408045.
PR   21-DEC-1989;   89US-0456142.
PR   17-SEP-1990;   90EP-0310149.
XX
PA   (CHIR ) CHIRON CORP.
PA   (OYAA/) OYA A.
XX
PI   Cha T,  Han J,  Houghton M,  Irvine BD,  Kolberg JA;
PI   Miyamura T,  Saito I,  Weiner AJ;
XX
DR   WPI; 1999-460843/41.
DR   P-PSDB; AAY14975.
XX
PT   New Hepatitis C Virus isolates, useful for diagnosis of hepatitis
PT   infections and development of vaccines
XX
PS   Disclosure; Fig 12; 132pp; English.
XX
CC   The invention provides two new isolates of hepatitis C virus (HCV), J1
CC   and J7. These two isolates comprise nucleotide and amino acid sequences
CC   that are distinct from the HCV isolate HCV-1. The nucleotide sequences
CC   may be used to detect non-A, non-B HCV (NANBH) polynucleotides by
CC   hybridisation for diagnosis of NANBH infections. They may also be used to
CC   screen blood donors, donated blood and blood products for this infection.
CC   The isolates may also be used to isolate other naturally occurring
CC   variants of the virus. The polypeptides may be used as a vaccine for
CC   administration to patients to protect against infection with NANBH. The
CC   present sequence represents the nucleotide sequence of HCV-1 ORF.
CC   (Updated on 20-MAR-2003 to correct PF field.)
CC   (Updated on 20-MAR-2003 to correct PR field.)
XX
SQ   Sequence 9133 BP; 1834 A; 2772 C; 2600 G; 1927 T; 0 other;

Alignment Scores:
Pred. No.:              1.74e-77     Length:        9133
Score:                  943.00       Matches:       179
Percent Similarity:     100.00%      Conservative:  3
Best Local Similarity:  98.35%       Mismatches:    0
Query Match:            98.95%       Indels:        0
DB:                     20           Gaps:          0

US-09-965-594-1 (1-182) x AAZ07656 (1-9133)
```

```
Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
Db       3343 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 3402

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db       3403 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 3462

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
Db       3463 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 3522

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db       3523 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 3582

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db       3583 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 3642

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
Db       3643 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 3702

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db       3703 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 3762

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db       3763 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 3822

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db       3823 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 3882

Qy        181 ArgSer 182
Db       3883 AGGTCC 3888
```

RESULT 12
AAQ05956

```
ID   AAQ05956 standard; DNA; 9185 BP.
XX
AC   AAQ05956;
XX
DT   25-MAR-2003  (updated)
DT   23-JAN-1991  (first entry)
XX
DE   Sense strand of the compiled Hepatitis C virus cDNA sequence.
XX
KW   Hepatitis C virus (HCV); antiviral agent; ss.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             320..9185
FT                   /*tag= a
FT   misc_RNA        1..1667
FT                   /*tag= b
FT                   /note="epitope within this region is claimed"
FT   misc_RNA        8978..9185
FT                   /*tag= c
FT                   /note="encodes an epitope that is claimed"
XX
PN   EP388232-A.
XX
PD   19-SEP-1990.
XX
PF   16-MAR-1990;   90EP-0302866.
XX
PR   18-MAY-1989;   89US-0355002.
PR   17-MAR-1989;   89US-0325338.
PR   20-APR-1989;   89US-0341334.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Houghton M,  Choo QL,  Kuo G;
XX
DR   WPI; 1990-284416/38.
DR   P-PSDB; AAR08124.
XX
PT   Hepatitis C virus DNA - used for producing probes,
PT   polypeptide(s), antibodies and anti-sense polynucleotide(s) for
PT   diagnosis and therapy.
XX
PS   Disclosure; Fig 17; 83pp; English.
XX
CC   HCV cDNA libraries were constructed using pooled serum from a
CC   chimpanzee with chronic HCV infection. A lambda gt11 library was
CC   screened with probes derived from previously isolated clones. The
CC   ORF is derived from the overlapping clones b114a, ag30a, CA205a,
CC   CA290a, CA216a, pi4a, CA167b, CA156e, CA84a, CA59a, K9-1, 26j, 13i,
CC   12f, 14i, 11b, 7f, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c,
CC   14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g ,15e, b5a and 16jh. These
CC   clones extend the sequence of the HCV genome reported in EP-318216.
CC   The upstream region from nucleotides -319 to +1348 (=1-1667 in this
CC   file) is covered by clones b114a, 18g, ag30a, CA205a, CA290a,
CC   CA216a, pi4a, CA167b, CA156e, CA84a and CA59a; nucleotides
CC   8659-8866 (=8978-9185 in this file) are covered by clones b5a and
CC   16jh.
CC   See also AAQ05955.
CC   (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 9185 BP; 1849 A; 2790 C; 2608 G; 1938 T; 0 other;
```

```
Alignment Scores:
Pred. No.:              1.75e-77      Length:         9185
Score:                  943.00        Matches:        179
Percent Similarity:     100.00%       Conservative:   3
Best Local Similarity:  98.35%        Mismatches:     0
Query Match:            98.95%        Indels:         0
DB:                     11            Gaps:           0
```

US-09-965-594-1 (1-182) x AAQ05956 (1-9185)

```
Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
Db       3395 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 3454

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db       3455 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 3514

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
Db       3515 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 3574

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db       3575 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 3634

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db       3635 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 3694

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
Db       3695 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 3754

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db       3755 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 3814

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
```

```
Db       3815 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 3874

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180

Db       3875 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 3934

Qy        181 ArgSer 182

Db       3935 AGGTCC 3940

RESULT 13
AAQ10566
ID   AAQ10566 standard; DNA; 9185 BP.
XX
AC   AAQ10566;
XX
DT   25-MAR-2003  (updated)
DT   29-APR-1991  (first entry)
XX
DE   Hepatitis C virus strain 1 DNA.
XX
KW   Hepatitis C virus; HCV-1; non-A, non-B hepatitis; HCV antigen;
KW   viral infections; ss.
XX
OS   Hepatitis C virus.
XX
PN   EP414475-A.
XX
PD   27-FEB-1991.
XX
PF   21-AUG-1990;   90EP-0309120.
XX
PR   25-AUG-1989;   89US-0398667.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Weiner AJ,  Steimer KS;
XX
DR   WPI; 1991-059670/09.
XX
PT   Cell lines infected with hepatitis C virus - are used as source
PT   of antigens for detection of HCV antibodies, for vaccines, and
PT   for screening anti-viral agents
XX
PS   Disclosure; fig 1; 24pp; English.
XX
CC   This is a hepatitis C virus (HCV) composite cDNA sequence, deduced
CC   using overlapping clones.  a compsn. contg. the antigenic protein
CC   encoded by this sequence is useful for detecting anti-HCV anti-
CC   bodies (Abs) and for screening an agent which inhibits HCV replic-
CC   ation.  A cell line infected with this virus can be used as a
CC   source of antigens.  The antigen is useful for preparing vaccines
CC   for treating viral infections.  See also AAQ10567.
CC   (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 9185 BP; 1849 A; 2790 C; 2608 G; 1938 T; 0 other;

Alignment Scores:
Pred. No.:              1.75e-77        Length:         9185
Score:                  943.00          Matches:        179
Percent Similarity:     100.00%         Conservative:   3
Best Local Similarity:  98.35%          Mismatches:     0
Query Match:            98.95%          Indels:         0
DB:                     12              Gaps:           0

US-09-965-594-1 (1-182) x AAQ10566 (1-9185)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20

Db       3395 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 3454

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40

Db       3455 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 3514

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60

Db       3515 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 3574

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80

Db       3575 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 3634

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100

Db       3635 CAAGACCTTGTGGGCTGGCCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 3694

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120

Db       3695 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 3754

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140

Db       3755 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 3814

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160

Db       3815 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 3874

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180

Db       3875 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 3934

Qy        181 ArgSer 182

Db       3935 AGGTCC 3940

RESULT 14
AAA75297
ID   AAA75297 standard; cDNA; 9185 BP.
XX
AC   AAA75297;
XX
DT   15-JAN-2001  (first entry)
XX
DE   Sense strand of HCV encoding a polyprotein.
XX
KW   Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
KW   viral infectivity; viral replication; ds.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             320..9184
FT                   /*tag= a
FT                   /note= "partial sequence; no termination codon given"
XX
PN   EP1034785-A2.
XX
PD   13-SEP-2000.
XX
PF   16-MAR-1990; 2000EP-0109602.
XX
PR   17-MAR-1989;   89US-0325338.
PR   20-APR-1989;   89US-0341334.
PR   18-MAY-1989;   89US-0355002.
PR   16-MAR-1990;   90EP-0302866.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Houghton M,  Choo Q,  Kuo G;
XX
DR   WPI; 2000-566891/53.
DR   P-PSDB; AAB18541.
XX
```

```
PT   Novel composition comprising a hepatitis C virus antisense
PT   polynucleotide which is complementary to or corresponds to a sense
PT   strand of the virus genome, and selectively hybridises to it -
XX
PS   Example; Fig 17; 75pp; English.
XX
CC   The specification describes a pharmaceutical composition which
CC   comprises a hepatitis C virus (HCV) antisense polynucleotide. The
CC   HCV is characterized by a positive stranded RNA genome which has
CC   40% homology at the polypeptide level to a HCV polyprotein. The
CC   antisense polynucleotide binds to cellular polynucleotides which
CC   enhance and/or are required for viral intectivity, replicative
CC   ability or chronicity. The antisense polynucleotides may also be
CC   designed to bind with high specificity, to be of increased stability,
CC   to be stable and to have low toxicity. The composition also comprises
CC   an agent which causes viral RNA to be inactive. The composition
CC   is used for preventing HCV replication in a system. The present
CC   sequence represents a novel HCV cDNA sequence, which is used in the
CC   course of the invention.
XX
SQ   Sequence 9185 BP; 1849 A; 2790 C; 2608 G; 1938 T; 0 other;

Alignment Scores:
Pred. No.:              1.75e-77        Length:         9185
Score:                  943.00          Matches:        179
Percent Similarity:     100.00%         Conservative:   3
Best Local Similarity:  98.35%          Mismatches:     0
Query Match:            98.95%          Indels:         0
DB:                     21              Gaps:           0

US-09-965-594-1 (1-182) x AAA75297 (1-9185)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              :::|||||:|||||| |||||:|||||.| ||||: |||:' |||| |||| ||||: ||||:
Db       3395 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 3454

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              |||||||:|||||||:|||||||||||||||:|||||:|||| ||||:'|||'||||||.|
Db       3455 AGCCTAACTGGCCGGCACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 3514

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              :|:||||:|:||||||||||||||||| |'||||. ||||:|.||||:|||'|||||
Db       3515 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 3574

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              |:|||||||||||||:|||||:|||||| ||||| |||':.||||:|||||| |||.|
Db       3575 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 3634

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              :::.||||||||||||:||||.|.||||| |||||..||||:|||| :||||| |||| '
Db       3635 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 3694

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              :'||||.|:||||:|.|||||'||||:| |||| | |||' |||:| ||| |||| :|
Db       3695 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGCGGG 3754

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              |.|||| '|||:|.||||, |||||| ||||'|.|||: ||||||||||| ||||. |
Db       3755 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 3814

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal Cys 160
              ||||:||||||.|||||'.||||:|||||||.||||'||||| ||||.|||| ||||.'|
Db       3815 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 3874

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              |||'|'||||||.|||'|||| ||||||:|||| :|||| |:::|'|||  |||| ||
Db       3875 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGACACAACCATG 3934

Qy        181 ArgSer 182
              ||'| |
Db       3935 AGGTCC 3940

RESULT 15

AAQ21744
ID   AAQ21744 standard; DNA; 9400 BP.
XX
AC   AAQ21744;
XX
DT   25-MAR-2003  (updated)
DT   22-JUN-1992  (first entry)
XX
DE   Compiled HCV cDNA.
XX
KW   HCV1; serum; qt11; oligomer; target; ds.
XX
OS   Hepatitis C virus 1.
XX
FH   Key             Location/Qualifiers
FT   CDS             341..9373
FT                   /*tag= a
FT                   /label= polyprotein
FT   misc_feature    357..386
FT                   /*tag= b
FT                   /label= target_region
FT                   /note= "see CC"
FT   misc_feature    390..419
FT                   /*tag= c
FT                   /label= target_region
FT                   /note= "see CC"
FT   misc_feature    423..452
FT                   /*tag= d
FT                   /label= target_region
FT                   /note= "see CC"
FT   misc_feature    456..485
FT                   /*tag= e
FT                   /label= target_region
FT                   /note= "see CC"
FT   misc_feature    489..518
FT                   /*tag= f
FT                   /label= target_region
FT                   /note= "see CC"
FT   misc_feature    552..581
FT                   /*tag= g
FT                   /label= target_region
FT                   /note= "see CC"
FT   misc_feature    583..612
FT                   /*tag= h
FT                   /label= target_region
FT                   /note= "see CC"
FT   misc_feature    615..645
FT                   /*tag= i
FT                   /label= target_region
FT                   /note= "see CC"
FT   misc_feature    673..702
FT                   /*tag= j
FT                   /label= target_region
FT                   /note= "see CC"
FT   misc_feature    706..735
FT                   /*tag= k
FT                   /label= target_region
FT                   /note= "see CC"
FT   misc_feature    739..768
FT                   /*tag= l
FT                   /label= target_region
FT                   /note= "see CC"
FT   misc_feature    798..827
FT                   /*tag= m
FT                   /label= target_region
FT                   /note= "see CC"
XX
PN   WO9202642-A.
XX
PD   20-FEB-1992.
XX
PF   12-AUG-1991;   91WO-US05728.
XX
```

```
PR   10-AUG-1990;   90US-0566209.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Houghton M,  Choo QL,  Kuo G,  Weiner AJ,  Urdea MS,  Irvine BD;
PI   Kolberg JA;
XX
DR   WPI; 1992-080094/10.
DR   P-PSDB; AAR21519.
XX
PT   Reagents for isolating, amplifying and detecting HCV
PT   polynucleotide(s) - used to monitor spread of blood-borne non-A,
PT   non-B hepatitis virus infection and screen blood samples for
PT   virus
XX
PS   Disclosure; Fig 1; 67pp; English.
XX
CC   The sequence is a composite of HCV cDNA from HCV1, a prototypic
CC   HCV.  The sequence is based upon sequence information derived from
CC   a no. of HCV cDNA clones, which were isolated form a no. of HCV
CC   cDNA libraries, including the "c" library present in lambda gt11
CC   (ATCC No.40394), and from human serum.  The HCV cDNA clones
CC   were isolated by methods described in WO9014436.
CC   The clones from which the sequence was derived are 5'clone32,
CC   b114a, 18g, ag30a, CA205a, CA290a, CA216a, pi14a, CA167b, CA156e,
CC   CA84a, CA59a, K9-1 (also called k9-1), 26j, 13i, 12f, 14i, 11b, 7f,
CC   7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g,
CC   39c, 35f, 19g, 26g, 15e, b5a, 16jh, 6k and pl31jh.
CC   The target regions indicated in the features are listed in claim 1,
CC   page 49.  Oligomers are provided which are complementary to these
CC   target regions and used in the detection of an HCV sequence in an
CC   analyte.
CC   (Updated on 25-MAR-2003 to correct PA field.)
CC   (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ   Sequence 9400 BP; 1885 A; 2860 C; 2671 G; 1964 T; 0 other;

Alignment Scores:
Pred. No.:              1.81e-77       Length:         9400
Score:                  943.00         Matches:        179
Percent Similarity:     100.00%        Conservative:   3
Best Local Similarity:  98.35%         Mismatches:     0
Query Match:            98.95%         Indels:         0
DB:                     13             Gaps:           0

US-09-965-594-1 (1-182) x AAQ21744 (1-9400)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              :::| ||||||| ||||.|.|||| ||||.||||| |||.. |||| |||:| |||
Db       3416 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 3475

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              ||.:||||||.||||:||||.:||||| |||| : |||| |||||| ||||: |||
Db       3476 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 3535

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              |||:|||||:|||||:|||||:||||||:.||||:.||||| ||||:.:||| : |||
Db       3536 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 3595

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              .|||||:||||||||||||.|:|||||||:|||||.|||:|||||:|||||||:|||
Db       3596 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 3655

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              :::|||||||||||||||||||||||||.||||.|||||||||||||:||||||||:|||
Db       3656 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 3715

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              |||||||:||||||||||||.||||||||||||||||.|||||||||:||||||| ||
Db       3716 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 3775

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              |:|||||||||||||:||||||||:|||||||||||||| ||||:||||||:||||| ||
Db       3776 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 3835

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              |||:||||:: |||||||||||||||||||||||||||||||||||||||||||:||||||
Db       3836 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 3895

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              ||:|||||||||:||||||.|||||| |||||| ||||.| |||:::||||||:||||:||
Db       3896 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 3955

Qy        181 ArgSer 182
              ||||::
Db       3956 AGGTCC 3961

Search completed: September 27, 2003, 12:36:32
Job time : 344 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 27, 2003, 12:22:50 ; Search time 3348 Seconds
(without alignments)
2223.881 Million cell updates/sec

Title: US-09-965-594-1
Perfect score: 953
Sequence: 1 MAPITAYAQQTRGLLGCIIT.....GVAKAVDFIPVESLETTMRS 182

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2668711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 99%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09965594/runat_26092003_164343_18756/app_query.fasta_1.327
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=99.9 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594_6CGN_1_1_3508_$runat_26092003_164344_18756 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP 6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	946	99.3	2058	6	AX395309	AX395309 Sequence
2	946	99.3	2058	6	AX454818	AX454818 Sequence
3	945	99.2	543	14	AF369218	AF369218 Hepatitis
4	945	99.2	543	14	AF369235	AF369235 Hepatitis
5	943	99.0	5360	6	AR118686	AR118686 Sequence
6	943	99.0	5360	6	I06434	I06434 Sequence 48
7	943	99.0	5360	6	I09328	I09328 Sequence 8
8	943	99.0	6785	6	AR118692	AR118692 Sequence
9	943	99.0	6785	6	I06440	I06440 Sequence 54
10	943	99.0	6785	6	I09329	I09329 Sequence 10
11	943	99.0	7310	6	AR118696	AR118696 Sequence
12	943	99.0	7310	6	I09331	I09331 Sequence 15
13	943	99.0	7310	14	HPCPOLYP	M32084 Hepatitis C
14	943	99.0	8316	6	AR118703	AR118703 Sequence
15	943	99.0	8987	6	AR118728	AR118728 Sequence
16	943	99.0	9185	6	AR118722	AR118722 Sequence
c 17	943	99.0	9185	6	AR118723	AR118723 Sequence
18	943	99.0	9185	6	BD091382	BD091382 HCV culti
19	943	99.0	9185	6	I08294	I08294 Sequence 1
20	943	99.0	9379	6	AR166930	AR166930 Sequence
21	943	99.0	9379	6	AR301300	AR301300 Sequence
22	943	99.0	9401	6	AR176483	AR176483 Sequence
23	943	99.0	9401	6	BD080334	BD080334 Hepatitis
24	943	99.0	9401	6	E66593	E66593 Hepatitis C
25	943	99.0	9401	6	I71894	I71894 Sequence 9
26	943	99.0	9401	6	I81885	I81885 Sequence 9
27	943	99.0	9401	14	HPCPLYPRE	M62321 Hepatitis C
28	943	99.0	9609	12	AF387805	AF387805 Synthetic
29	943	99.0	9609	12	AF387808	AF387808 Synthetic
30	943	99.0	9618	14	AF271632	AF271632 Hepatitis
31	943	99.0	9646	12	AF387806	AF387806 Synthetic
32	943	99.0	9693	12	AF387807	AF387807 Synthetic
33	942	98.8	543	14	AF369222	AF369222 Hepatitis
34	942	98.8	543	14	AF369232	AF369232 Hepatitis
35	942	98.8	543	14	AF369240	AF369240 Hepatitis
36	942	98.8	543	14	AF369245	AF369245 Hepatitis
37	942	98.8	2061	6	AX441176	AX441176 Sequence
38	942	98.8	2061	6	AX467113	AX467113 Sequence
39	940	98.6	543	14	AF369224	AF369224 Hepatitis
40	940	98.6	543	14	AF369237	AF369237 Hepatitis
41	940	98.6	9424	14	AF511948	AF511948 Hepatitis
42	939	98.5	543	14	AF369230	AF369230 Hepatitis
43	939	98.5	9502	6	E08263	E08263 gRNA of Hep
44	939	98.5	9502	6	E08264	E08264 cDNA of Hep
45	939	98.5	9502	14	HPCHCJ1	D10749 Hepatitis C

ALIGNMENTS

RESULT 1

```
AX395309
LOCUS       AX395309                2058 bp    DNA     linear   PAT 18-MAY-2002
DEFINITION  Sequence 2 from Patent WO0196875.
ACCESSION   AX395309
VERSION     AX395309.1  GI:21066308
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Chien,D.Y., Arcangel,P., Tandeske,L., George-Nascimento,C.,
            Coit,D. and Medina-Selby,A.
  TITLE     Hcv antigen/antibody combination assay
  JOURNAL   Patent: WO 0196875-A 2 20-DEC-2001;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2058
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="representative NS3/4a conformational antigen"
     CDS             1. .>2058
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAD38234.1"
                     /db_xref="GI:21066309"
                     /translation="MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTF
                     LATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGS
                     SDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVC
                     TRGVAKAVDFIPVENLETTMRSPVFTDNSSPPVVPQSFQVAHLHAPTGSGKSTKVPAA
                     YAAQGYKVLVLNPSVAATLGFGAYMSKAHGIDPNIRTGVRTITTGSPITYSTYGKFLA
                     DGGCSGGAYDIIICDECHSTDATSILGIGTVLDQAETAGARLVVLATATPPGSVTVPH
                     PNIEEVALSTTGEIPFYGKAIPLEVIKGGRHLIFCHSKKKCDELAAKLVALGINAVAY
                     YRGLDVSVIPPIGDVVVVATDALMTGYTGDFDSVIDCNTCVTQTVDFSLDPTFTIETT
                     TLPQDAVSRTQRRGRTGRGKPGIYRFVAPGERPSGMFDSSVLCECYDAGCAWYELTPA
                     ETTVRLRAYMNTPGLPVCQDHLEFWEGVFTGLTHIDAHFLSQTKQSGENLPYLVAYQA
                     TVCARAQAPPPSWDQMWKCLIRLKPTLHGPTPLLYRLGAVQNEITLTHPVTKYIMTCM
                     SADLEVVTSTWVLVGGVLAALAAYCLSTGCVVIVGRVVLSGKPAIIPDREVLYREFDE
                     MEEC"
BASE COUNT      419 a    634 c    580 g    425 t
ORIGIN

Alignment Scores:
Pred. No.:              6.56e-68        Length:         2058
Score:                  946.00          Matches:        180
Percent Similarity:     100.00%         Conservative:   2
Best Local Similarity:  98.90%          Mismatches:     0
Query Match:            99.27%          Indels:         0
DB:                     6               Gaps:           0

US-09-965-594-1 (1-182) x AX395309 (1-2058)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              ||||| |||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 60

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCA 120

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCAAACGTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCA 180

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 240

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGATCATTGACACCCTGCACTTGC 300

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATCCCGGTGCGCCGGCGG 360

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 420

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 480

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 540

Qy        181 ArgSer 182
              ||||||
Db        541 AGGTCC 546

RESULT 2
AX454818
LOCUS       AX454818                2058 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 1 from Patent WO0196870.
ACCESSION   AX454818
VERSION     AX454818.1  GI:21714047
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Chien,D.Y., Arcangel,P., Tandeske,L., George-Nascimento,C.,
            Coit,D. and Medina-Selby,A.
  TITLE     Immunoassays for anti-hcv antibodies
  JOURNAL   Patent: WO 0196870-A 1 20-DEC-2001;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2058
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="representative NS3/4a conformational antigen"
     CDS             1. .>2058
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAD38232.1"
                     /db_xref="GI:21714048"
                     /translation="MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTF
                     LATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGS
                     SDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVC
                     TRGVAKAVDFIPVENLETTMRSPVFTDNSSPPVVPQSFQVAHLHAPTGSGKSTKVPAA
                     YAAQGYKVLVLNPSVAATLGFGAYMSKAHGIDPNIRTGVRTITTGSPITYSTYGKFLA
                     DGGCSGGAYDIIICDECHSTDATSILGIGTVLDQAETAGARLVVLATATPPGSVTVPH
                     PNIEEVALSTTGEIPFYGKAIPLEVIKGGRHLIFCHSKKKCDELAAKLVALGINAVAY
                     YRGLDVSVIPPIGDVVVVATDALMTGYTGDFDSVIDCNTCVTQTVDFSLDPTFTIETI
                     TLPQDAVSRTQRRGRTGRGKPGIYRFVAPGERPSGMFDSSVLCECYDAGCAWYELTPA
                     ETTVRLRAYMNTPGLPVCQDHLEFWEGVFTGLTHIDAHFLSQTKQSGENLPYLVAYQA
                     TVCARAQAPPPSWDQMWKCLIRLKPTLHGPTPLLYRLGAVQNEITLTHPVTKYIMTCM
                     SADLEVVTSTWVLVGGVLAALAAYCLSTGCVVIVGRVVLSGKPAIIPDREVLYREFDE
                     MEEC"
BASE COUNT      419 a    633 c    581 g    425 t
ORIGIN

Alignment Scores:
Pred. No.:              6.56e-68        Length:         2058
Score:                  946.00          Matches:        180
Percent Similarity:     100.00%         Conservative:   2
Best Local Similarity:  98.90%          Mismatches:     0
Query Match:            99.27%          Indels:         0
DB:                     6               Gaps:           0

US-09-965-594-1 (1-182) x AX454818 (1-2058)
```

```
Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
Db          1 ATGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTGCTAGGGTGCATAATCACC 60

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db         61 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 120

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
Db        121 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 180

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db        181 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 240

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db        241 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGATCATTGACACCCTGCACTTGC 300

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
Db        301 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 360

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db        361 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 420

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db        421 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 480

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db        481 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 540

Qy        181 ArgSer 182
Db        541 AGGTCC 546
```

```
RESULT 3
AF369218
LOCUS       AF369218                 543 bp    RNA     linear   VRL 03-JUL-2002
DEFINITION  Hepatitis C virus Pt.4 NS3 protease gene, partial cds.
ACCESSION   AF369218
VERSION     AF369218.1  GI:14150560
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Holland-Staley,C.A., Kovari,L.C., Golenberg,E.M., Pobursky,K.J. and
            Mayers,D.L.
  TITLE     Genetic diversity and response to IFN of the NS3 protease gene from
            clinical strains of the hepatitis C virus
  JOURNAL   Arch. Virol. 147 (7), 1385-1406 (2002)
  MEDLINE   22105140
   PUBMED   12111414
REFERENCE   2  (bases 1 to 543)
  AUTHORS   Holland-Staley,C.A., Kovari,L.C., Golenberg,E. and Mayers,D.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2001) Infectious Disease Research, Henry Ford
            Health Systems, 2799 W. Grand Blvd. Rm 7045 E & R, Detroit, MI
            48202, USA
FEATURES             Location/Qualifiers
     source          1..543
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /isolate="Pt.4"
                     /db_xref="taxon:11103"
                     /note="type: 1A"
     CDS             <1..>543
                     /note="polyprotein"
                     /codon_start=1
                     /product="NS3 protease"
                     /protein_id="AAK54543.1"
                     /db_xref="GI:14150561"
                     /translation="APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFL
                     ATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSS
                     DLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCT
                     RGVAKAVDFIPVENLETTMRS"
BASE COUNT      113 a    165 c    158 g    106 t
ORIGIN

Alignment Scores:
Pred. No.:              1.89e-68        Length:         543
Score:                  945.00          Matches:        180
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.45%          Mismatches:     0
Query Match:            99.16%          Indels:         0
DB:                     14              Gaps:           0

US-09-965-594-1 (1-182) x AF369218 (1-543)
```

```
Qy          2 AlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThrSer 21
Db          1 GCACCTATCACGGCATACGCCCAGCAGACAAGGGGCCTCCTAGGGTGTATAATCACCAGT 60

Qy         22 LeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAla 41
Db         61 CTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCTGCC 120

Qy         42 GlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGly 61
Db        121 CAGACTTTCTTGGCAACATGCATCAACGGGGTATGCTGGACCGTCTACCACGGGGCCGGA 180

Qy         62 ThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAspLys 81
Db        181 ACGAGGACCATTGCATCACCCAAAGGTCCCGTCATCCAGATGTACACCAATGTAGACAAA 240

Qy         82 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCysGly 101
Db        241 GATCTTGTGGGCTGGCCCGCTCCTCAAGGTTCCCGCTCATTGACACCCTGCACCTGCGGC 300

Qy        102 SerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArgGly 121
Db        301 TCCTCGGACCTTTACTTGGTCACGAGGCATGCCGATGTCATACCCGTGCGCCGGCGAGGT 360

Qy        122 AspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGly 141
Db        361 GATAGCAGGGGCAGCCTGCTCTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCGGGG 420

Qy        142 GlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThr 161
Db        421 GGTCCGCTGTTGTGCCCCGCGGGACACGCCGTAGGCATATTCAGGGCCGCGGTGTGCACC 480

Qy        162 ArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMetArg 181
Db        481 CGTGGAGTGGCTAAGGCGGTGGACTTTATCCCCGTGGAGAACCTAGAGACAACCATGAGG 540

Qy        182 Ser 182
Db        541 TCC 543
```

```
RESULT 4
AF369235
LOCUS       AF369235                 543 bp    RNA     linear   VRL 03-JUL-2002
DEFINITION  Hepatitis C virus Pt.1Y NS3 protease gene, partial cds.
ACCESSION   AF369235
VERSION     AF369235.1  GI:14150594
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
```

```
          Hepacivirus.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Holland-Staley,C.A., Kovari,L.C., Golenberg,E.M., Pobursky,K.J. and
            Mayers,D.L.
  TITLE     Genetic diversity and response to IFN of the NS3 protease gene from
            clinical strains of the hepatitis C virus
  JOURNAL   Arch. Virol. 147 (7), 1385-1406 (2002)
  MEDLINE   22105140
   PUBMED   12111414
REFERENCE   2  (bases 1 to 543)
  AUTHORS   Holland-Staley,C.A., Kovari,L.C., Golenberg,E. and Mayers,D.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2001) Infectious Disease Research, Henry Ford
            Health Systems, 2799 W. Grand Blvd. Rm 7045 E & R, Detroit, MI
            48202, USA
FEATURES             Location/Qualifiers
     source          1..543
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /isolate="PT.1Y"
                     /db_xref="taxon:11103"
                     /note="type: 1A"
     CDS             <1..>543
                     /note="polyprotein"
                     /codon_start=1
                     /product="NS3 protease"
                     /protein_id="AAK54560.1"
                     /db_xref="GI:14150595"
                     /translation="APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTF
                     ATCINGVCWTVYHGAGTRTIASPKGPVTQMYTNVDKDLVGWPAPQGSRSLTPCTCGSS
                     DLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCT
                     RGVAKAVDFIPVENLETTMRS"
BASE COUNT      112 a    169 c    159 g    103 t
ORIGIN

Alignment Scores:
Pred. No.:                1.89e-68        Length:          543
Score:                    945.00          Matches:         180
Percent Similarity:       100.00%         Conservative:    1
Best Local Similarity:    99.45%          Mismatches:      0
Query Match:              99.16%          Indels:          0
DB:                       14              Gaps:            0

US-09-965-594-1 (1-182) x AF369235 (1-543)

Qy          2 AlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThrSer 21
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCACCCATCACAGCATACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACCAGC 60

Qy         22 LeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAla 41
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTAACCGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCTGCC 120

Qy         42 GlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGly 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAAACTTCCTTGGCAACGTGCATCAATGGGGTATGCTGGACCGTCTACCACGGGGCCGGA 180

Qy         62 ThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAspLys 81
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACGAGGACCATCGCATCACCAAAGGGTCCCGTTATCCAGATGTACACCAATGTAGACAAA 240

Qy         82 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCysGly 101
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GATCTTGTGGGCTGGCCCGCTCCTCAAGGTTCCCGCTCATTGACACCCTGCACCTGCGGC 300

Qy        102 SerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArgGly 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGGGGT 360

Qy        122 AspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGly 141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GATAGCAGGGGCAGCCTGCTCTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCGGGG 420

Qy        142 GlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThr 161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGTCCGCTGTTGTGCCCCCGCGGGACACGCCGTAGGCATATTCAGGGCCGCGGTGTGCACC 480

Qy        162 ArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMetArg 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATGAGG 540

Qy        182 Ser 182
              |||
Db        541 TCC 543

RESULT 5
AR118686
LOCUS       AR118686               5360 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 53 from patent US 6150087.
ACCESSION   AR118686
VERSION     AR118686.1  GI:14100596
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5360)
  AUTHORS   Chien,D.Y.
  TITLE     NANBV diagnostics and vaccines
  JOURNAL   Patent: US 6150087-A 53 21-NOV-2000;
FEATURES             Location/Qualifiers
     source          1..5360
                     /organism="unknown"
BASE COUNT     1060 a   1623 c   1532 g   1145 t
ORIGIN

Alignment Scores:
Pred. No.:                3.22e-67        Length:          5360
Score:                    943.00          Matches:         179
Percent Similarity:       100.00%         Conservative:    3
Best Local Similarity:    98.35%          Mismatches:      0
Query Match:              98.95%          Indels:          0
DB:                       6               Gaps:            0

US-09-965-594-1 (1-182) x AR118686 (1-5360)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        930 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 989

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        990 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1049

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1050 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 1109

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1110 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 1169

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1170 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 1229

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1230 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 1289

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1290 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 1349

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1350 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 1409
```

```
Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1410 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 1469

Qy        181 ArgSer 182
              ||||||
Db       1470 AGGTCC 1475

RESULT 6
I06434
LOCUS       I06434                  5360 bp    DNA     linear   PAT 02-DEC-1994
DEFINITION  Sequence 48 from Patent EP 0318216.
ACCESSION   I06434
VERSION     I06434.1  GI:590311
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5360)
  AUTHORS   Houghton,M., Choo,Q.-L. and Kuo,G.
  TITLE     Nanbv diagnostics and vaccines
  JOURNAL   Patent: EP 0318216-A1 48 31-MAY-1989;
FEATURES             Location/Qualifiers
     source          1. .5360
                     /organism="unknown"
BASE COUNT     1061 a   1623 c   1533 g   1143 t
ORIGIN

Alignment Scores:
Pred. No.:                3.22e-67        Length:          5360
Score:                    943.00          Matches:         179
Percent Similarity:       100.00%         Conservative:    3
Best Local Similarity:    98.35%          Mismatches:      0
Query Match:              98.95%          Indels:          0
DB:                       6               Gaps:            0

US-09-965-594-1 (1-182) x I06434 (1-5360)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        930 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 989

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        990 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1049

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1050 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 1109

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1110 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 1169

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1170 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 1229

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1230 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 1289

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1290 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 1349

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1350 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 1409

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1410 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 1469

Qy        181 ArgSer 182
              ||||||
Db       1470 AGGTCC 1475

RESULT 7
I09328
LOCUS       I09328                  5360 bp    DNA     linear   PAT 02-DEC-1994
DEFINITION  Sequence 8 from Patent WO 8904669.
ACCESSION   I09328
VERSION     I09328.1  GI:587963
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5360)
  AUTHORS   Houghton,M., Choo,Q.-K. and Kuo,G.
  JOURNAL   Patent: WO 8904669-A 8 01-JUN-1989;
FEATURES             Location/Qualifiers
     source          1. .5360
                     /organism="unknown"
BASE COUNT     1061 a   1623 c   1533 g   1143 t
ORIGIN

Alignment Scores:
Pred. No.:                3.22e-67        Length:          5360
Score:                    943.00          Matches:         179
Percent Similarity:       100.00%         Conservative:    3
Best Local Similarity:    98.35%          Mismatches:      0
Query Match:              98.95%          Indels:          0
DB:                       6               Gaps:            0

US-09-965-594-1 (1-182) x I09328 (1-5360)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        930 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 989

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        990 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1049

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1050 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 1109

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1110 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 1169

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1170 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 1229

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1230 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 1289

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1290 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 1349

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1350 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 1409

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1410 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 1469

Qy        181 ArgSer 182
              ||||||
```

```
Db       1470 AGGTCC 1475

RESULT 8
AR118692
LOCUS       AR118692                6785 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 65 from patent US 6150087.
ACCESSION   AR118692
VERSION     AR118692.1  GI:14100602
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6785)
  AUTHORS   Chien,D.Y.
  TITLE     NANBV diagnostics and vaccines
  JOURNAL   Patent: US 6150087-A 65 21-NOV-2000;
FEATURES             Location/Qualifiers
     source          1..6785
                     /organism="unknown"
BASE COUNT     1392 a   2050 c   1914 g   1429 t
ORIGIN

Alignment Scores:
Pred. No.:                4.15e-67      Length:          6785
Score:                    943.00        Matches:         179
Percent Similarity:       100.00%       Conservative:    3
Best Local Similarity:    98.35%        Mismatches:      0
Query Match:              98.95%        Indels:          0
DB:                       6             Gaps:            0

US-09-965-594-1 (1-182) x AR118692 (1-6785)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1203 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 1262

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1263 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1322

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1323 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 1382

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1383 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 1442

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1443 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 1502

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1503 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 1562

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1563 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 1622

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1623 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTCGGCATATTTAGGGCCGCGGTGTGC 1682

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1683 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 1742

Qy        181 ArgSer 182
              ||||||
Db       1743 AGGTCC 1748

RESULT 9
I06440
LOCUS       I06440                  6785 bp    DNA     linear   PAT 02-DEC-1994
DEFINITION  Sequence 54 from Patent EP 0318216.
ACCESSION   I06440
VERSION     I06440.1  GI:590312
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6785)
  AUTHORS   Houghton,M., Choo,Q.-L. and Kuo,G.
  TITLE     Nanbv diagnostics and vaccines
  JOURNAL   Patent: EP 0318216-A1 54 31-MAY-1989;
FEATURES             Location/Qualifiers
     source          1..6785
                     /organism="unknown"
BASE COUNT     1392 a   2050 c   1914 g   1429 t
ORIGIN

Alignment Scores:
Pred. No.:                4.15e-67      Length:          6785
Score:                    943.00        Matches:         179
Percent Similarity:       100.00%       Conservative:    3
Best Local Similarity:    98.35%        Mismatches:      0
Query Match:              98.95%        Indels:          0
DB:                       6             Gaps:            0

US-09-965-594-1 (1-182) x I06440 (1-6785)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1203 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 1262

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1263 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1322

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1323 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 1382

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1383 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 1442

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1443 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 1502

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1503 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 1562

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1563 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 1622

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1623 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTCGGCATATTTAGGGCCGCGGTGTGC 1682

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1683 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 1742

Qy        181 ArgSer 182
              ||||||
Db       1743 AGGTCC 1748

RESULT 10
I09329
LOCUS       I09329                  6785 bp    DNA     linear   PAT 02-DEC-1994
DEFINITION  Sequence 10 from Patent WO 8904669.
```

```
ACCESSION   I09329
VERSION     I09329.1  GI:587964
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6785)
  AUTHORS   Houghton,M., Choo,Q.-K. and Kuo,G.
  JOURNAL   Patent: WO 8904669-A 10 01-JUN-1989;
FEATURES             Location/Qualifiers
     source          1..6785
                     /organism="unknown"
BASE COUNT     1392 a   2050 c   1914 g   1429 t
ORIGIN

Alignment Scores:
Pred. No.:                4.15e-67       Length:          6785
Score:                    943.00         Matches:         179
Percent Similarity:       100.00%        Conservative:    3
Best Local Similarity:    98.35%         Mismatches:      0
Query Match:              98.95%         Indels:          0
DB:                       6              Gaps:            0

US-09-965-594-1 (1-182) x I09329 (1-6785)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1203 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 1262

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1263 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1322

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1323 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 1382

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1383 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 1442

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1443 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 1502

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1503 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 1562

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1563 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 1622

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1623 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 1682

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1683 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 1742

Qy        181 ArgSer 182
              ||||||
Db       1743 AGGTCC 1748

RESULT 11
AR118696
LOCUS       AR118696                7310 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 74 from patent US 6150087.
ACCESSION   AR118696
VERSION     AR118696.1  GI:14100606
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7310)
  AUTHORS   Chien,D.Y.
  TITLE     NANBV diagnostics and vaccines
  JOURNAL   Patent: US 6150087-A 74 21-NOV-2000;
FEATURES             Location/Qualifiers
     source          1..7310
                     /organism="unknown"
BASE COUNT     1495 a   2220 c   2056 g   1539 t
ORIGIN

Alignment Scores:
Pred. No.:                4.5e-67        Length:          7310
Score:                    943.00         Matches:         179
Percent Similarity:       100.00%        Conservative:    3
Best Local Similarity:    98.35%         Mismatches:      0
Query Match:              98.95%         Indels:          0
DB:                       6              Gaps:            0

US-09-965-594-1 (1-182) x AR118696 (1-7310)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1728 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 1787

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1788 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1847

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1848 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 1907

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1908 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 1967

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1968 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 2027

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2028 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 2087

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2088 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 2147

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2148 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 2207

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2208 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 2267

Qy        181 ArgSer 182
              ||||||
Db       2266 AGGTCC 2273

RESULT 12
I09331
LOCUS       I09331                  7310 bp    DNA     linear   PAT 02-DEC-1994
DEFINITION  Sequence 15 from Patent WO 8904669.
ACCESSION   I09331
VERSION     I09331.1  GI:587966
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7310)
```

```
  AUTHORS   Houghton,M., Choo,Q.-K. and Kuo,G.
  JOURNAL   Patent: WO 8904669-A 15 01-JUN-1989;
FEATURES             Location/Qualifiers
     source          1. .7310
                     /organism="unknown"
BASE COUNT     1495 a    2218 c    2058 g    1539 t
ORIGIN

Alignment Scores:
Pred. No.:                4.5e-67        Length:          7310
Score:                    943.00         Matches:         179
Percent Similarity:       100.00%        Conservative:    3
Best Local Similarity:    98.35%         Mismatches:      0
Query Match:              98.95%         Indels:          0
DB:                       6              Gaps:            0

US-09-965-594-1 (1-182) x IO9331 (1-7310)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1728 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCTCCTAGGGTGCATAATCACC 1787

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1788 AGCCTAACTGGCCGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1847

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1848 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 1907

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1908 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 1967

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1968 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 2027

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2028 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 2087

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2088 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 2147

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2148 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 2207

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2208 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 2267

Qy        181 ArgSer 182
              ||||||
Db       2268 AGGTCC 2273

RESULT 13
HPCPOLYP
LOCUS       HPCPOLYP                7310 bp ss-RNA     linear   VRL 02-AUG-1993
DEFINITION  Hepatitis C virus polyprotein gene, partial cds.
ACCESSION   M32084
VERSION     M32084.1  GI:329875
KEYWORDS    polyprotein.
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 7310)
  AUTHORS   Choo,Q.-L., Richman,K. and Han,J.
  TITLE     The nucleotide sequence of the Hepatitis C viral genome
  JOURNAL   Unpublished (1990)
COMMENT     Original source text: Hepatitis C virus, cDNA to viral RNA, clones
            K9-1 through 15e, isolated from chimpanzee (individual 910) blood
            plasma.
            Draft entry and printed sequence for [1] kindly submitted by
            M.Houghton, 22-FEB-1990. Chiron Corporation, 4560 Horton Street,
            Emeryville CA 94608.
FEATURES             Location/Qualifiers
     source          1. .7310
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:11103"
     CDS             <1. .>7310
                     /note="polyprotein"
                     /codon_start=3
                     /protein_id="AAA45677.1"
                     /db_xref="GI:329876"
                     /translation="GCPERLASCRPLPDPDQGWGPISYANGSGPDQRPYCWHYPPKPC
                     GIVPAKSVCGPVYCFTPSPVVVGTIDRSGAPIYSWGENDTDVFVLNNTRPPLGNWFGC
                     TWMNSTGFTKVCGAPPCVIGGAGNNILHCPTDCFRKHPDATYSRCGSGPWITPRCLVD
                     YPYRLWHYPCTINYTIFKIRMYVGGVEHRLEAACNWTRGERCDLEDRDRSELSPLLLT
                     TTQWQVLPCSFTTLPALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVVLLFLLL
                     ADARVCSCLWMMLLISQAEAALENLVILNAASLAGTHGLVSFLVFFCFAWYLKGKWVP
                     GAVYTFYGMWPLLLLLLALPQRAYALDTEVAASCGGVVLVGLMALTLSPYYKRYISWC
                     LWWLQYFLTRVEAQLHVWIPPLNVRGGRDAVILLMCAVHPTLVFDITKLLLAVFGPLW
                     ILQASLLKVPYFVRVQGLLRFCALARKMIGGHYVQMVIIKLGALTGTYVYNHLTPLRD
                     WAHNGLRDLAVAVEPVVFSQMETKLITWGADTAACGDIINGLPVSARRGREILLGPAD
                     GMVSKGWRLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCI
                     NGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYL
                     VTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVA
                     KAVDFIPVENLETTMRSPVFTDNSSPPVVPQSFQVAHLHAPTGSGKSTKVPAAYAAQG
                     YKVLVLNPSVAATLGFGAYMSKAHGIDPNIRTGVRTITIGSPITYSTYGKFLADGGCS
                     GGAYDIIICDECHSTDATSILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEE
                     VALSTTGEIPFYGKAIPLEVIKGGRHLIFCHSKKKCDELAAKLVALGINAVAYYRGLD
                     VSVIPTSGDVVVVATDALMTGYTGDFDSVIDCNTCVTQTVDFSLDPTFTIETITLPQD
                     AVSRTQRRGRTGRGKPGIYRFVAPGERPSGMFDSSVLCECYDAGCAWYELTPAETTVR
                     LRAYMNTPGLPVCQDHLEFWEGVFTGLTHIDAHFLSQTKQSGENLPYLVAYQATVCAR
                     AQAPPPSWDQMWKCLIRLKPTLHGPTPLLYRLGAVQNEITLTHPVTKYIMTCMSADLE
                     VVTSTWVLVGGVLAALAAYCLSTGCVVIVGRVVLSGKPAIIPDREVLYREFDEMEECS
                     QHLPYIEQGMMLAEQFKQKALGLLQTASRQAEVIAPAVQTNWQKLETFWAKHMWNFIS
                     GIQYLAGLSTLPGNPAIASLMAFTAAVTSPLTTSQTLLFNILGGWVAAQLAAPGAATA
                     FVGAGLAGAAIGSVGLGKVLIDILAGYGAGVAGALVAFKIMSGEVPSTEDLVNLLPAI
                     LSPGALVVGVVCAAILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARV
                     TAILSSLTVTQLLRRLHQWISSECTTPCSGSWLRDIWDWICEVLSDFKTWLKAKLMPQ
                     LPGIPFVSCQRGYKGVWRVDGIMHTRCHCGAEITGHVKNGTMRIVGPRTCRNMWSGTF
                     PINAYTTGPCTPLPAPNYTFALWRVSAEEYVEIRQVGDFHYVTGMTTDNLKCPCQVPS
                     PEFFTELDGVRLHRFAPPCKPLLREEVSFRVGLHEYPVGSQLPCEPEPDVAVLTSMLT
                     DPSHITAEAAGRRLARGSPPSVASSSASQLSAPSLKATCTANHDSPDAELIEANLLWR
                     QEMGGNITRVESENKVVILDSFDPLVAEEDEREISVPAEILRKSRRFAQALPVWARPD
                     YNPPLVETWKKPDYEPPVVHGCPLPPPKSPPVPPPRKKRTVVLTESTLSTALAELATR
                     SFGSSSTSGITGDNTTTSSEPAPSGCPPDSDAESYSSMPPLEGEPGDPDLSDGSWSTV
                     SSEANAEDVVCCSMSYSWTGALVTPCAAEEQKLPINALSNSLLRHHNLVYSTTSRSAC
                     QRQKKVTFDRLQVLDSHYQDVLKEVKAAASKVKANLLSVEEACSLTPPHSAKSKFGYG
                     AKDVRCHARKAVTHINSVWKDLLEDNVTPIDTTIMAKNEVFCVQPEKGGRKPARLIVF
                     PDLGVRVCEKMALYDVVTKLPLAVMGSSYGFQYSPGQRVEFLVQAWKSKKTPMGFSYD
                     TRCFDSTVTESDIRTEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGENCGYRR
                     CRASGVLTTSCGNTLTCYIKARAACRAAGLQDCTMLVCGDDLVVICESAGVQEDAASL
                     RAFTEAMTRYSAPPGDPPQPEYDLELITSCSSNVSVAHDGAGKRVYYLTRDPTTPLAR
                     AAWETARHTPVNSWLGNIIMFAPTLWARMILMTHFFSVLIARDQLEQALDCEIYGACY
                     SIEPLDLPPIIQRL"
BASE COUNT     1495 a    2218 c    2058 g    1539 t
ORIGIN

Alignment Scores:
Pred. No.:                4.5e-67        Length:          7310
Score:                    943.00         Matches:         179
Percent Similarity:       100.00%        Conservative:    3
Best Local Similarity:    98.35%         Mismatches:      0
Query Match:              98.95%         Indels:          0
DB:                       14             Gaps:            0

US-09-965-594-1 (1-182) x HPCPOLYP (1-7310)

Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
```

```
Db       1728 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 1787

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db       1788 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1847

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
Db       1848 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 1907

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db       1908 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 1967

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db       1968 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 2027

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
Db       2028 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 2087

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db       2088 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 2147

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db       2148 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 2207

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db       2208 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 2267

Qy        181 ArgSer 182
Db       2268 AGGTCC 2273
```

RESULT 14
AR118703
LOCUS AR118703 8316 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 88 from patent US 6150087.
ACCESSION AR118703
VERSION AR118703.1 GI:14100613
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 8316)
AUTHORS Chien,D.Y.
TITLE NANBV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 88 21-NOV-2000;
FEATURES Location/Qualifiers
source 1..8316
/organism="unknown"
BASE COUNT 1671 a 2529 c 2545 g 1771 t
ORIGIN

Alignment Scores:

Pred. No.:	5.17e-67	Length:	8316
Score:	943.00	Matches:	179
Percent Similarity:	100.00%	Conservative:	3
Best Local Similarity:	98.35%	Mismatches:	0
Query Match:	98.95%	Indels:	0
DB:	6	Gaps:	0

US-09-965-594-1 (1-182) x AR118703 (1-8316)

```
Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
Db       2734 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 2793

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db       2794 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 2853

Qy         41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
Db       2854 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 2913

Qy         61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db       2914 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 2973

Qy         81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db       2974 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 3033

Qy        101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
Db       3034 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 3093

Qy        121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db       3094 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTTCCTACTTGAAAGGCTCCTCG 3153

Qy        141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db       3154 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 3213

Qy        161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db       3214 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 3273

Qy        181 ArgSer 182
Db       3274 AGGTCC 3279
```

RESULT 15
AR118728
LOCUS AR118728 8987 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 137 from patent US 6150087.
ACCESSION AR118728
VERSION AR118728.1 GI:14100638
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 8987)
AUTHORS Chien,D.Y.
TITLE NANBV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 137 21-NOV-2000;
FEATURES Location/Qualifiers
source 1..8987
/organism="unknown"
BASE COUNT 1807 a 2735 c 2547 g 1898 t
ORIGIN

Alignment Scores:

Pred. No.:	5.62e-67	Length:	8987
Score:	943.00	Matches:	179
Percent Similarity:	100.00%	Conservative:	3
Best Local Similarity:	98.35%	Mismatches:	0
Query Match:	98.95%	Indels:	0
DB:	6	Gaps:	0

US-09-965-594-1 (1-182) x AR118728 (1-8987)

```
Qy          1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
Db       3076 CTGGCGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATCACC 3135

Qy         21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db       3136 AGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 3195
```

```
Qy        41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
             ||||||||:|:||||||.|.||||:|||||:|:|||.||||||:|,||||:|||:.||||:|
Db      3196 GCCCAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 3255

Qy        61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
             ||||||||||:||||:|||||:|:||||| |||:|||||||.||||| ||| |||||| |
Db      3256 GGAACGAGGACCATCGCGTCACCCAAGGGTCCTGTCATCCAGATGTATACCAATGTAGAC 3315

Qy        81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
             :::|||||:|||| |:|||:| ||||:|||| :|||| .||||.|||||:|||.:|
Db      3316 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGACACCCTGCACTTGC 3375

Qy       101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
             |.||||:|||||:|.||||:|||:|.|||:|'|||' ||||| ||| .:|||: ||
Db      3376 GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGCGCCGGCGG 3435

Qy       121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
             ||||:| ||||:| ||||:| |||| |.|||' |||| '||||. |||' ||| :.||
Db      3436 GGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCCATTCCTACTTGAAAGGCTCCTCG 3495

Qy       141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIleProArgAlaAlaValCys 160
             ||||:|||||:| ||||:|:|||| |.|||' |||: |||:. ||||. |||:  |||
Db      3496 GGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 3555

Qy       161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
               |||||| |||||:|||| |||| |||,' |||:  |||:::|||:| ||| ||||
Db      3556 ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 3615

Qy       181 ArgSer 182
             ||||||
Db      3616 AGGTCC 3621
```

Search completed: September 27, 2003, 13:32:26
Job time : 3361 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 27, 2003, 11:35:53 ; Search time 107 Seconds
(without alignments)
436.931 Million cell updates/sec

Title: US-09-965-594-1
Perfect score: 953
Sequence: 1 MAPITAYAQQTRGLLGCIIT.....GVAKAVDFIPVESLETTMRS 182

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 256052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 99%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	945	99.2	181	12	Q91RR8	Q91rr8 hepatitis c
2	945	99.2	181	12	Q91RT5	Q91rt5 hepatitis c
3	943	99.0	2436	12	Q81756	Q81756 hepatitis c
4	943	99.0	3011	12	Q9IFE5	Q9ife5 hepatitis c
5	942	98.8	181	12	Q91RR3	Q91rr3 hepatitis c
6	942	98.8	181	12	Q91RS1	Q91rs1 hepatitis c
7	942	98.8	181	12	Q91RQ8	Q91rq8 hepatitis c
8	942	98.8	181	12	Q91RT1	Q91rt1 hepatitis c
9	940	98.6	181	12	Q91RR6	Q91rr6 hepatitis c
10	940	98.6	181	12	Q91RS9	Q91rs9 hepatitis c
11	939	98.5	181	12	Q91RS3	Q91rs3 hepatitis c
12	939	98.5	3011	12	Q03463	Q03463 hepatitis c
13	938	98.4	181	12	Q91RT4	Q91rt4 hepatitis c
14	938	98.4	181	12	Q91RS8	Q91rs8 hepatitis c
15	938	98.4	181	12	Q91RT3	Q91rt3 hepatitis c
16	938	98.4	181	12	Q91RS5	Q91rs5 hepatitis c
17	938	98.4	181	12	Q91RS7	Q91rs7 hepatitis c
18	938	98.4	181	12	Q91RT0	Q91rt0 hepatitis c
19	937	98.3	3011	12	O36579	O36579 hepatitis c
20	936	98.2	181	12	Q91RR5	Q91rr5 hepatitis c
21	936	98.2	181	12	Q91RS4	Q91rs4 hepatitis c
22	935	98.1	181	12	Q91RT6	Q91rt6 hepatitis c
23	934	98.0	181	12	Q91RT9	Q91rt9 hepatitis c
24	933	97.9	181	12	Q91RR4	Q91rr4 hepatitis c
25	933	97.9	181	12	Q91RR9	Q91rr9 hepatitis c
26	933	97.9	181	12	Q91RR0	Q91rr0 hepatitis c
27	933	97.9	3011	12	Q9ELS8	Q9els8 hepatitis c
28	932	97.8	181	12	Q91RR2	Q91rr2 hepatitis c
29	931	97.7	181	12	Q91RR7	Q91rr7 hepatitis c
30	931	97.7	3011	12	Q9DIT6	Q9dit6 hepatitis c
31	930	97.6	181	12	Q91RT2	Q91rt2 hepatitis c
32	930	97.6	181	12	Q91RR1	Q91rr1 hepatitis c
33	930	97.6	181	12	Q91RQ9	Q91rq9 hepatitis c
34	930	97.6	181	12	Q91RS2	Q91rs2 hepatitis c
35	930	97.6	3011	12	O36608	O36608 hepatitis c
36	930	97.6	3015	12	Q9PWX5	Q9pwx5 hepatitis c
37	930	97.6	3015	12	Q9PWU9	Q9pwu9 hepatitis c
38	928	97.4	181	12	Q91RS6	Q91rs6 hepatitis c
39	927	97.3	181	12	Q91RT7	Q91rt7 hepatitis c
40	925	97.1	181	12	Q91RS0	Q91rs0 hepatitis c
41	925	97.1	181	12	Q91RT8	Q91rt8 hepatitis c
42	924	97.0	3011	12	O36609	O36609 hepatitis c
43	919	96.4	3011	12	O36610	O36610 hepatitis c
44	909	95.4	3011	12	Q81754	Q81754 hepatitis c
45	903	94.8	3010	12	Q9J3G9	Q9j3g9 hepatitis c

ALIGNMENTS

RESULT 1
Q91RR8
ID Q91RR8 PRELIMINARY; PRT; 181 AA.
AC Q91RR8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pt.1Y;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF369235; AAK54560.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19130 MW; 85D9186929987C35 CRC64;

Query Match 99.2%; Score 945; DB 12; Length 181;
Best Local Similarity 99.4%; Pred. No. 1.9e-88;
Matches 180; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          2 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
              ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 60

Qy         62 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 120
```

```
Qy        122 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTMR 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTMR 180

Qy        182 S 182
              |
Db        181 S 181


RESULT 2
Q91RT5
ID   Q91RT5      PRELIMINARY;      PRT;   181 AA.
AC   Q91RT5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   NS3 protease (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=P1.4;
RA   Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT   "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT   Clinical Strains of the Hepatitis C Virus.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF369218; AAK54543.1; -.
DR   InterPro; IPR004109; HCV_NS3.
DR   Pfam; PF02907; HCV_NS3; 1.
KW   Protease.
FT   NON_TER       1       1
FT   NON_TER     181     181
SQ   SEQUENCE   181 AA;  19130 MW;  85D9186B299B7C35 CRC64;

  Query Match             99.2%;  Score 945;  DB 12;  Length 181;
  Best Local Similarity   99.4%;  Pred. No. 1.9e-88;
  Matches  180;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
              ||.|||||||||||||||.||||.||||.|||||.||||.|||||.||||.||||.||||
Db          1 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 60

Qy         62 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 120

Qy        122 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTMR 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTMR 180

Qy        182 S 182
              |
Db        181 S 181


RESULT 3
Q81756
ID   Q81756      PRELIMINARY;      PRT;  2436 AA.
AC   Q81756;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Genome polyprotein (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Choo Q.-L., Richman K., Han J.;
RT   "The nucleotide sequence of the Hepatitis C viral genome.";
RL   Submitted (MAY-1990) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; M32084; AAA45677.1; -.
DR   HSSP; P27958; 1A1V.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS50507; RDRP_POSITIVE; 1.
DR   PROSITE; PS50521; RDRP_VIRAL; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
FT   NON_TER       1       1
FT   NON_TER    2436    2436
SQ   SEQUENCE   2436 AA;  264734 MW;  D7B9872900BE3125 CRC64;

  Query Match             99.0%;  Score 943;  DB 12;  Length 2436;
  Best Local Similarity   98.4%;  Pred. No. 7.1e-87;
  Matches  179;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :||.|||||||||||||.||||.||||.|||||.||||.|||||.||||.||||.||||
Db        576 LAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 635

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              |:|||||||||||||.||||:|||||||||||||.||||:|:||||||||||||||||||
Db        636 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 695

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              |||||||:|||||||||||||||||||||||||||||||:|||||||||||||||:|:|||
Db        696 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 755

Qy        181 RS 182
              ||
Db        756 RS 757


RESULT 4
Q9IFE5
ID   Q9IFE5      PRELIMINARY;      PRT;  3011 AA.
AC   Q9IFE5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21262212; PubMed=11369872;
RA   Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT   "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
RT   sequence.";
RL   J. Gen. Virol. 82:1291-1297(2001).
```

```
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF271632; AAF81759.1; -.
DR   HSSP; P27958; 1A1V.
DR   InterPro; IPR000345; CytC_heme_bind.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01550; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
DR   PROSITE; PS50507; RDRP_POSITIVE; 1.
DR   PROSITE; PS50521; RDRP_VIRAL; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3011 AA;  327124 MW;  2489CE74AC864E58 CRC64;

  Query Match             99.0%;  Score 943;  DB 12;  Length 3011;
  Best Local Similarity   98.4%;  Pred. No. 9.2e-87;
  Matches  179;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 LAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETIM 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETIM 1205

Qy        181 RS 182
              ||
Db       1206 RS 1207


RESULT 5
Q91RR3
ID   Q91RR3      PRELIMINARY;      PRT;   181 AA.
AC   Q91RR3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   NS3 protease (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pt.4B;
RA   Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT   "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT   Clinical Strains of the Hepatitis C Virus.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF369240; AAK54565.1; -.
DR   InterPro; IPR004109; HCV_NS3.
DR   Pfam; PF02907; HCV_NS3; 1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     181    181
SQ   SEQUENCE   181 AA;  19115 MW;  5D85F88AD7AC1A11 CRC64;

  Query Match             98.8%;  Score 942;  DB 12;  Length 181;
  Best Local Similarity   98.9%;  Pred. No. 3.8e-88;
  Matches  179;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 60

Qy         62 TRIIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
              |||||||||||:|||||||:||||||||||:||||||||||||||||||||||||||||||
Db         61 TRIIASPKGPVIQMYTNVDRDLVGWPAPQGARSLTPCTCGSSDLYLVTRHADVIPVRRRG 120

Qy        122 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTMR 181
              |||||||||||||.|||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTMR 180

Qy        182 S 182
              |
Db        181 S 181


RESULT 6
Q91RS1
ID   Q91RS1      PRELIMINARY;      PRT;   181 AA.
AC   Q91RS1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   NS3 protease (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pt.K;
RA   Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT   "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT   Clinical Strains of the Hepatitis C Virus.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF369232; AAK54557.1; -.
DR   InterPro; IPR004109; HCV_NS3.
DR   Pfam; PF02907; HCV_NS3; 1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     181    181
SQ   SEQUENCE   181 AA;  19114 MW;  ABB90B5B3ABA4E26 CRC64;

  Query Match             98.8%;  Score 942;  DB 12;  Length 181;
  Best Local Similarity   98.9%;  Pred. No. 3.8e-88;
  Matches  179;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 60
```

```
Qy         62 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
              ||||| ||| ||||||||||||.||||| |||| |||| ||||| |||| |||| |
Db         61 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGARSLTPCTCGSSDLYLVTRHADVIPVRRRG 120

Qy        122 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTMR 181
              |||||.|||||||||||.|||||||||||||.||||| |||| |||||||||||||||:|| |||
Db        121 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTMR 180

Qy        182 S 182
              |
Db        181 S 181


RESULT 7
Q91RQ8
ID   Q91RQ8      PRELIMINARY;      PRT;   181 AA.
AC   Q91RQ8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   NS3 protease (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pt.5Z;
RA   Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT   "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT   Clinical Strains of the Hepatitis C Virus.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF369245; AAK54570.1; -.
DR   InterPro; IPR004109; HCV_NS3.
DR   Pfam; PF02907; HCV_NS3; 1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     181    181
SQ   SEQUENCE   181 AA;  19144 MW;  C0C91F1E2EEB0B32 CRC64;

  Query Match             98.8%;  Score 942;  DB 12;  Length 181;
  Best Local Similarity   98.9%;  Pred. No. 3.8e-88;
  Matches  179;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db          1 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 60

Qy         62 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGTRSLTPCTCGSSDLYLVTRHADVIPVRRRG 120

Qy        122 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTMR 181
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTMR 180

Qy        182 S 182
              |
Db        181 S 181


RESULT 8
Q91RT1
ID   Q91RT1      PRELIMINARY;      PRT;   181 AA.
AC   Q91RT1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   NS3 protease (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pt.16I;
RA   Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT   "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT   Clinical Strains of the Hepatitis C Virus.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF369222; AAK54547.1; -.
DR   InterPro; IPR004109; HCV_NS3.
DR   Pfam; PF02907; HCV_NS3; 1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     181    181
SQ   SEQUENCE   181 AA;  19114 MW;  ABB90B5B3ABA4E26 CRC64;

  Query Match             98.8%;  Score 942;  DB 12;  Length 181;
  Best Local Similarity   98.9%;  Pred. No. 3.8e-88;
  Matches  179;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db          1 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 60

Qy         62 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGARSLTPCTCGSSDLYLVTRHADVIPVRRRG 120

Qy        122 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTMR 181
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTMR 180

Qy        182 S 182
              |
Db        181 S 181


RESULT 9
Q91RR6
ID   Q91RR6      PRELIMINARY;      PRT;   181 AA.
AC   Q91RR6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   NS3 protease (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pt.3T;
RA   Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT   "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT   Clinical Strains of the Hepatitis C Virus.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF369237; AAK54562.1; -.
DR   InterPro; IPR004109; HCV_NS3.
DR   Pfam; PF02907; HCV_NS3; 1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     181    181
SQ   SEQUENCE   181 AA;  19101 MW;  614ADA8B0F33CCAF CRC64;

  Query Match             98.6%;  Score 940;  DB 12;  Length 181;
  Best Local Similarity   98.3%;  Pred. No. 6e-88;
  Matches  178;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
              |||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db          1 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEIQIVSTAAQTFLATCINGVCWTVYHGAG 60

Qy         62 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
              |||||||||||||||||:||||||||:||||||||||||||||||||||||||||||||||
```

```
Db         61 TRTIASPKGPVIQMYTNVDQDLVGWPAPQGARSLTPCTCGSSDLYLVTRHADVIPVRRRG 120

Qy        122 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTMR 181
              ||:||||||:|:|||||: |||| |:|||:  ||||:  ||| |:||| :||| |:||
Db        121 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTMR 180

Qy        182 S 182
              |
Db        181 S 181


RESULT 10
Q91RS9
ID   Q91RS9      PRELIMINARY;      PRT;   181 AA.
AC   Q91RS9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   NS3 protease (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pt.174;
RA   Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT   "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT   Clinical Strains of the Hepatitis C Virus.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF369224; AAK54549.1; -.
DR   InterPro; IPR004109; HCV_NS3.
DR   Pfam; PF02907; HCV_NS3; 1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     181    181
SQ   SEQUENCE   181 AA;  19131 MW;  8BD7FC2769DBD635 CRC64;

  Query Match             98.6%;  Score 940;  DB 12;  Length 181;
  Best Local Similarity   98.9%;  Pred. No. 6e-88;
  Matches  179;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
              :||||||:|.|||||:||||| |||:|| |||| :|||: ||||| ||| :|||
Db          1 APITAYAQQTRGLLGCIITSLTGRDKDQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 60

Qy         62 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
              ||||||:|||||:||||||||:||||| |||| :||||| ||||  ||||: |||:
Db         61 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 120

Qy        122 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTMR 181
              ||||||.:|||||.|.|||||.||||:.||||:: |||:  ||||  :||||: ||||
Db        121 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTMR 180

Qy        182 S 182
              |
Db        181 S 181


RESULT 11
Q91RS3
ID   Q91RS3      PRELIMINARY;      PRT;   181 AA.
AC   Q91RS3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   NS3 protease (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pt.24;
RA   Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT   "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT   Clinical Strains of the Hepatitis C Virus.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF369230; AAK54555.1; -.
DR   InterPro; IPR004109; HCV_NS3.
DR   Pfam; PF02907; HCV_NS3; 1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     181    181
SQ   SEQUENCE   181 AA;  19132 MW;  0BB90B5F3AB95250 CRC64;

  Query Match             98.5%;  Score 939;  DB 12;  Length 181;
  Best Local Similarity   98.3%;  Pred. No. 7.6e-88;
  Matches  178;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
              ||||||||:| ||||:||||||||||||||| |||||||||||||||||||||||||||||
Db          1 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 60

Qy         62 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
              |||:|:|||||||||||||:||||||||||:|||||||||:||||||||:|||||||||||
Db         61 TRTMASPKGPVIQMYTNVDKDLVGWPAPQGARSLTPCTCGSSDLYLVTRHADVIPVRRRG 120

Qy        122 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTMR 181
              |||||.||||||||||||||||||||||||||||||||||:||||||:|||||:||||||
Db        121 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTMR 180

Qy        182 S 182
              :
Db        181 S 181


RESULT 12
Q03463
ID   Q03463      PRELIMINARY;      PRT;  3011 AA.
AC   Q03463;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HC-J1;
RX   MEDLINE=91013116; PubMed=2170712;
RA   Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA   Yoshizawa H.;
RT   "The 5'-terminal sequence of the hepatitis C virus genome.";
RL   Jpn. J. Exp. Med. 60:167-177(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HC-J1;
RX   MEDLINE=92044440; PubMed=1658196;
RA   Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
RA   Miyakawa Y., Mayumi M.;
RT   "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT   from a human carrier: comparison with reported isolates for conserved
RT   and divergent regions.";
RL   J. Gen. Virol. 72:2697-2704(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HC-J1;
RX   MEDLINE=93117120; PubMed=1335573;
RA   Okamoto H., Kanai N., Mishiro S.;
RT   "Full-length nucleotide sequence of a Japanese hepatitis C virus
RT   isolate (HC-J1) with high homology to USA isolates.";
RL   Nucleic Acids Res. 20:6410-6410(1992).
RN   [4]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=HC-J1;
RA   Okamoto H.;
RL   Submitted (DEC-1992) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HC-J1;
RX   MEDLINE=94174722; PubMed=7510436;
RA   Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,
RA   Inchauspe G.;
RT   "Characterization and mapping of a B-cell immunogenic domain in
RT   hepatitis C virus E2 glycoprotein using a yeast peptide library.";
RL   Virology 200:246-255(1994).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; D10749; BAA01582.1; -.
DR   HSSP; P27958; 1HE1.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS50507; RDRP_POSITIVE; 1.
DR   PROSITE; PS50521; RDRP_VIRAL; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3011 AA;  327112 MW;  97E9052C02504635 CRC64;

  Query Match             98.5%;  Score 939;  DB 12;  Length 3011;
  Best Local Similarity   97.8%;  Pred. No. 2.4e-86;
  Matches  178;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 LAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              ||||||||||||||||||||:|||||||||||:||||||||||||||||||||||||||||
Db       1086 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGARSLTPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              |||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db       1146 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHVVGIFRAAVCTRGVAKAVDFIPVESLETTM 1205

Qy        181 RS 182
              ||
Db       1206 RS 1207
```

```
RESULT 13
Q91RT4
ID   Q91RT4      PRELIMINARY;      PRT;   181 AA.
AC   Q91RT4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   NS3 protease (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pt.23;
RA   Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT   "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT   Clinical Strains of the Hepatitis C Virus.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF369219; AAK54544.1; -.
DR   InterPro; IPR004109; HCV_NS3.
DR   Pfam; PF02907; HCV_NS3; 1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     181    181
SQ   SEQUENCE   181 AA;  19059 MW;  1E53C47AE8B7E5C9 CRC64;

  Query Match             98.4%;  Score 938;  DB 12;  Length 181;
  Best Local Similarity   98.3%;  Pred. No. 9.6e-88;
  Matches  178;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 60

Qy         62 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
              |:|||||||||||||||||:|||||||||||:|||||||||||||||||||||||||||||
Db         61 TKTIASPKGPVIQMYTNVDQDLVGWPAPQGARSLTPCTCGSSDLYLVTRHADVIPVRRRG 120

Qy        122 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTMR 181
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTMR 180

Qy        182 S 182
              |
Db        181 S 181


RESULT 14
Q91RS8
ID   Q91RS8      PRELIMINARY;      PRT;   181 AA.
AC   Q91RS8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   NS3 protease (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pt.176;
RA   Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT   "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT   Clinical Strains of the Hepatitis C Virus.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF369225; AAK54550.1; -.
DR   InterPro; IPR004109; HCV_NS3.
DR   Pfam; PF02907; HCV_NS3; 1.
KW   Protease.
FT   NON_TER       1      1
```

```
FT   NON_TER     181    181
SQ   SEQUENCE   181 AA;  19114 MW;  574AC47AE8AEE5D2 CRC64;

  Query Match             98.4%;  Score 938;  DB 12;  Length 181;
  Best Local Similarity   98.3%;  Pred. No. 9.6e-88;
  Matches  178;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
              |||||||||||||||| ||||||| |||| |||| | |||| |||| |||| ||||
Db          1 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 60

Qy         62 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TRTIASPKGPVIQMYTNVDQDLVGWPAPQGARSLTPCTCGSSDLYLVTRHADVIPVRRRG 120

Qy        122 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTMR 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTMR 180

Qy        182 S 182
              |
Db        181 S 181


RESULT 15
Q91RT3
ID   Q91RT3      PRELIMINARY;      PRT;   181 AA.
AC   Q91RT3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   NS3 protease (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pt.11;
RA   Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT   "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT   Clinical Strains of the Hepatitis C Virus.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF369220; AAK54545.1; -.
DR   InterPro; IPR004109; HCV_NS3.
DR   Pfam; PF02907; HCV_NS3; 1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     181    181
SQ   SEQUENCE   181 AA;  19116 MW;  9848807F43EBD043 CRC64;

  Query Match             98.4%;  Score 938;  DB 12;  Length 181;
  Best Local Similarity   98.3%;  Pred. No. 9.6e-88;
  Matches  178;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
              |||||||||||||||||||||| |||||| |||| | |||| |||| | ||||||||| |
Db          1 APITAYAQQTRGLLGCIITSLTGRDKNQAEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 60

Qy         62 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
              |||||||||||||| |||| | |||| ||||||||||| |||||||||| |||| |
Db         61 TRTIASPKGPVIQMYTNVDKDLVGWPAPQGTRSLTPCTCGSSDLYLVTRHADVIPVRRRG 120

Qy        122 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTMR 181
              ||||| |||||||||||||||||||||| |||||||||||| |||||||||| |||| |
Db        121 DSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTMR 180

Qy        182 S 182
              |
Db        181 S 181


Search completed: September 27, 2003, 12:21:21
```

```
Job time : 109 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 26, 2003, 22:20:58 ; Search time 78 Seconds
(without alignments)
370.362 Million cell updates/sec

Title: US-09-965-594-1
Perfect score: 953
Sequence: 1 MAPITAYAQQTRGLLGCIIT.....GVAKAVDFIPVESLETTMRS 182

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107862

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 99%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	946	99.3	686	23	AAE18689	HCV-1 NS3/4a mutan
2	946	99.3	686	23	AAU76377	Hepatitis C virus
3	946	99.3	686	24	ABG72261	HCV-1 NS3/4a confo
4	944	99.1	3011	14	AAR40120	HCV genomic amino
5	943	99.0	609	15	AAR51170	Hepatitis C virus
6	943	99.0	1766	10	AAP92041	Sequence encoded i
7	943	99.0	1786	10	AAP90158	Protein sequence o
8	943	99.0	2261	10	AAP90164	Peptide encoded by
9	943	99.0	2301	10	AAP92047	Sequence encoded i
10	943	99.0	2436	10	AAP92050	Sequence encoded i
11	943	99.0	2436	10	AAP90288	Peptide encoded by
12	943	99.0	2772	21	AAB18540	Protein encoded by
13	943	99.0	2816	14	AAR34009	HCV-1 polyprotein.
14	943	99.0	2894	16	AAR70230	Composite hepatiti
15	943	99.0	2955	20	AAY14975	Amino acid sequenc
16	943	99.0	2955	21	AAB18541	Polyprotein encode
17	943	99.0	3011	13	AAR21519	Compiled HCV seque
18	943	99.0	3011	14	AAR31621	Hepatitis C virus
19	943	99.0	3011	17	AAR90931	Hepatitis C virus
20	943	99.0	3011	18	AAW34480	HCV polyprotein.
21	943	99.0	3011	19	AAW40038	HCV polyprotein.
22	943	99.0	3011	23	AAE22049	Hepatitis C virus
23	943	99.0	3011	23	AAU84597	HCV polyprotein 1a
24	942	98.8	632	23	AAE21847	Hepatitis C virus
25	942	98.8	632	23	AAE19905	Hepatitis C virus
26	942	98.8	686	23	AAE21837	Hepatitis C virus
27	942	98.8	686	23	AAE21838	Hepatitis C virus
28	942	98.8	686	23	AAE21839	Hepatitis C virus
29	942	98.8	686	23	AAE21840	Hepatitis C virus
30	942	98.8	686	23	AAE21841	Hepatitis C virus
31	942	98.8	686	23	AAE21842	Hepatitis C virus
32	942	98.8	686	23	AAE21843	Hepatitis C virus
33	942	98.8	686	23	AAE21844	Hepatitis C virus
34	942	98.8	686	23	AAE21845	Hepatitis C virus
35	942	98.8	686	23	AAE21846	Hepatitis C virus
36	942	98.8	686	23	AAE19900	Hepatitis C virus
37	942	98.8	686	23	AAE19907	Hepatitis C virus
38	942	98.8	686	23	AAE19908	Hepatitis C virus
39	942	98.8	686	23	AAE19919	Hepatitis C virus
40	942	98.8	686	23	AAE19920	Hepatitis C virus
41	942	98.8	686	23	AAE19921	Hepatitis C virus
42	942	98.8	686	23	AAE19922	Hepatitis C virus
43	942	98.8	686	23	AAE19923	Hepatitis C virus
44	942	98.8	686	23	AAE19924	Hepatitis C virus
45	942	98.8	686	23	AAE19925	Hepatitis C virus

ALIGNMENTS

RESULT 1
AAE18689
ID AAE18689 standard; Protein; 686 AA.
XX
AC AAE18689;
XX
DT 17-MAY-2002 (first entry)
XX
DE HCV-1 NS3/4a mutant conformational antigen.
XX
KW Hepatitis C virus; NS3/4a antigen; HCV infection; mutant; mutein.
XX
OS Hepatitis C virus type 1.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 403
FT /note= "Wild type Thr substituted with Pro"
FT Misc-difference 404
FT /note= "Wild type Ser substituted with Ile"
XX
PN WO200196875-A2.
XX
PD 20-DEC-2001.
XX
PF 14-JUN-2001; 2001WO-US19369.
XX
PR 15-JUN-2000; 2000US-212082P.
PR 02-APR-2001; 2001US-280811P.
PR 02-APR-2001; 2001US-280867P.
XX
PA (CHIR) CHIRON CORP.

```
XX
PI   Chien DY,  Arcangel P,  Tandeske L,  George-Nascimento C,  Coit D;
PI   Medina-Selby A;
XX
DR   WPI; 2002-179522/23.
DR   N-PSDB; AAD29795.
XX
PT   Immunoassay solid support useful for detecting hepatitis C virus
PT   infection in a biological sample, comprises at least one of HCV
PT   anti-core antibody and HCV NS3/4a epitope, bound to the support -
XX
PS   Example 2; Fig 4; 87pp; English.
XX
CC   The present invention relates to hepatitis C virus (HCV) core antigen
CC   and NS (nonstructural) 3/4a antibody combination assay that can detect
CC   both HCV antigens and antibodies present in a sample using a single
CC   solid matrix as well as immunoassay solid supports for use in the assay.
CC   The solid support is useful for detecting HCV infection in a biological
CC   sample. The present sequence is HCV-1 NS3/4a mutant conformational
CC   antigen. This sequence is used in the exemplification of the invention.
XX
SQ   Sequence   686 AA;

  Query Match           99.3%;  Score 946;  DB 23;  Length 686;
  Best Local Similarity   98.9%;  Pred. No. 1.2e-89;
  Matches  180;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 180

Qy        181 RS 182
              ||
Db        181 RS 182


RESULT 2
AAU76377
ID   AAU76377 standard; Protein; 686 AA.
XX
AC   AAU76377;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Hepatitis C virus NS3/4a conformational epitope protein sequence.
XX
KW   Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion;
KW   immunoassay solid support; multiple epitope fusion antigen; MEFA;
KW   non-structural protein; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 403
FT                   /note  "Wild-type Thr substituted by Pro"
FT   Misc-difference 404
FT                   /note= "Wild-type Ser substituted by Ile"
XX
PN   WO200196870-A2.
XX
PD   20-DEC-2001.
XX
PF   14-JUN-2001; 2001WO-US19156.
XX
PR   15-JUN-2000; 2000US-212082P.
PR   02-APR-2001; 2001US-280811P.
PR   02-APR-2001; 2001US-280867P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Chien DY,  Arcangel P,  Tandeske L,  George-nasciemento C,  Coit D;
PI   Medina-selby A;
XX
DR   WPI; 2002-090228/12.
DR   N-PSDB; ABK15344.
XX
PT   Immunoassay solid support, useful for detecting hepatitis C virus
PT   infection in biological sample, comprises HCV NS3/4a conformational
PT   epitope and multiple epitope fusion antigen bound to the support -
XX
PS   Claim 5; Fig 3; 92pp; English.
XX
CC   The present invention relates to a new immunoassay solid support
CC   consisting essentially of at least one hepatitis C virus (HCV) NS3/4a
CC   conformational epitope and a multiple epitope fusion antigen (MEFA),
CC   bound to the support. The NS3/4a conformational epitope and/or
CC   MEFA reacts specifically with anti-HCV antibodies present in a biological
CC   sample from an HCV-infected individual. The immunoassay of the invention
CC   is useful for detecting hepatitis C virus infection in a biological
CC   sample. The method of the invention provides a sensitive, accurate
CC   diagnostic and prognostic tool to provide adequate patient care and to
CC   prevent transmission of HCV by blood and by blood products, or by
CC   personal contact. Use of NS3/4a conformational epitope in combination
CC   with MEFA, provides a sensitive and reliable method for detecting early
CC   HCV seroconversion. Use of MEFA has the added advantages of decreasing
CC   masking problems, improving sensitivity in detecting antibodies by
CC   allowing a greater number of epitopes on a unit surface area of
CC   substrate, and improving substrate. Detection accuracy is increased and
CC   the incidence of false results is reduced because of the identification
CC   and the use of highly immunogenic HCV antigens which are present during
CC   the early stages of HCV seroconversion. The present amino acid sequence
CC   represents the non-structural protein NS3/4a conformational epitope of
CC   the invention.
XX
SQ   Sequence   686 AA;

  Query Match           99.3%;  Score 946;  DB 23;  Length 686;
  Best Local Similarity   98.9%;  Pred. No. 1.2e-89;
  Matches  180;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 180

Qy        181 RS 182
              ||
Db        181 RS 182


RESULT 3
ABG72261
ID   ABG72261 standard; Protein; 686 AA.
XX
AC   ABG72261;
XX
DT   06-MAR-2003  (first entry)
XX
DE   HCV-1 NS3/4a conformational antigen.
XX
```

```
KW   Immunoassay solid support; Hepatitis C Virus type-1; HCV-1;
KW   NS3/4a conformational epitope; multiple epitope fusion antigen;
KW   MEFA; anti-HCV antibody; NS3/4a conformational antigen;
KW   HCV infection; mutant; mutein.
XX
OS   Hepatitis C virus type 1.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Region          2..686
FT                   /note= "Corresponds to amino acid residues 1027-1711
FT                   of HCV-1 NS3/4a polypeptide"
FT   Misc-difference 403
FT                   /note= "Substitution of wild-type Thr to Pro"
FT   Misc-difference 404
FT                   /note= "Substitution of wild-type Ser to Ile"
XX
PN   US20021466B5-A1.
XX
PD   10-OCT-2002.
XX
PF   14-JUN-2001; 2001US-0881654.
XX
PR   15-JUN-2000; 2000US-212082P.
PR   02-APR-2001; 2001US-280811P.
PR   02-APR-2001; 2001US-280867P.
XX
PA   (CHIE/) CHIEN D Y.
PA   (ARCA/) ARCANGEL P.
PA   (TAND/) TANDESKE L.
PA   (GEOR/) GEORGE-NASCIMENTO C.
PA   (COIT/) COIT D.
PA   (MEDI/) MEDINA-SELBY A.
XX
PI   Chien DY,  Arcangel P,  Tandeske L,  George-Nascimento C,  Coit D;
PI   Medina-Selby A;
XX
DR   WPI; 2003-147573/14.
DR   N-PSDB; ABX14410.
XX
PT   Immunoassay solid support for detecting Hepatitis C Virus infection in
PT   biological samples, comprises Hepatitis C Virus conformational epitope
PT   and multiple epitope fusion antigen -
XX
PS   Claim 2; Fig 3A-3D; 45pp; English.
XX
CC   The present invention relates to immunoassays comprising Hepatitis C
CC   Virus (HCV) NS3/4a conformational epitope and multiple epitope fusion
CC   antigen (MEFA), bound to a solid support. The NS3/4a epitope and/or
CC   the multiple epitope fusion antigen react with anti-HCV antibodies
CC   present in a biological sample from an HCV-infected individual. The
CC   immunoassays and methods of the invention are useful for detecting
CC   HCV infection in a biological sample. The inventive immunoassay solid
CC   support provides a sensitive and reliable method for detecting early
CC   HCV seroconversion. The assays can detect HCV infection caused by any
CC   six known genotypes of HCV. The use of the multiple epitope fusion
CC   proteins decreases masking problems, improves sensitivity in detecting
CC   antibodies by allowing a greater number of epitopes on a unit area
CC   of substrate, and improves selectivity. The present sequence
CC   represents HCV type 1 (HCV-1) NS3/4a conformational antigen, a mutant
CC   of the HCV-1 NS3/4a polypeptide.
XX
SQ   Sequence   686 AA;

  Query Match             99.3%;  Score 946;  DB 24;  Length 686;
  Best Local Similarity   98.9%;  Pred. No. 1.2e-89;
  Matches  180;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 180

Qy        181 RS 182
              ||
Db        181 RS 182
```

```
RESULT 4
AAR40120
ID   AAR40120 standard; Protein; 3011 AA.
XX
AC   AAR40120;
XX
DT   25-MAR-2003  (updated)
DT   27-JAN-1994  (first entry)
XX
DE   HCV genomic amino acid sequence isolated from infected human LG.
XX
KW   Hepatitis C Virus; Non-A, non-B hepatitis Virus; HCV; NANBHV;
KW   human growth hormone; HGH; secretion signal; fusion protein;
KW   vaccine.
XX
OS   Hepatitis C Virus.
XX
PN   WO9315193-A1.
XX
PD   05-AUG-1993.
XX
PF   29-JAN-1993;   93WO-US00907.
XX
PR   31-JAN-1992;   92US-0830024.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PT   Bode SL,  Casey JM,  Desai SM,  Devare SG,  Frail DE;
PT   Yamaguchi J,  Zeck BJ;
XX
DR   WPI; 1993-258673/32.
XX
PT   New plasmid pHCV-162 is a mammalian expression systems for HCVD
PT   proteins - useful for diagnosing HCV infection and as vaccines
PT   for preventing HCV infection
XX
PS   Example 1; Page 39-49; 100pp; English.
XX
CC   RNA was isolated from the plasma of a HCV seropositive human
CC   (designated "LG") and cDNA was prepared from it. The cDNA was
CC   PCR amplified using specific primers with sequences based
CC   on the prototype HCV-1 cDNA sequence (GENBANK M62321). Further
CC   amplification using nested primers resulted in 7 adjacent HCV DNA
CC   fragments which could be assembled into a full-length sequence. The
CC   DNA sequence was determined and translated into the genomic amino
CC   acid sequence. Comparison of the LG genomic amino acid sequence
CC   with that from HCV-1 showed 134 amino acid differences.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   3011 AA;

  Query Match             99.1%;  Score 944;  DB 14;  Length 3011;
  Best Local Similarity   98.4%;  Pred. No. 1.4e-88;
  Matches  179;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 LAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              ||||||||||||||||||||:|||||||||:||:|||||||||||||||||||||||||||
```

```
Db       1086 GTRTIASPKGPVIQMYTNVDRDLVGWPAPQGARSLTPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              |||||||.|||||.| ||||.|||||||.||| |||| ||||||.||| |||||. |||
Db       1146 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 1205

Qy        181 RS 182
              ||
Db       1206 RS 1207


RESULT 5
AAR51170
ID   AAR51170 standard; peptide; 609 AA.
XX
AC   AAR51170;
XX
DT   20-OCT-1994  (first entry)
XX
DE   Hepatitis C virus non-structural protein 3.
XX
KW   Peptide; antibody; hepatitis C virus; HCV; identification;
KW   diagnosis; non-A non-B hepatitis; NANB; detection.
XX
OS   Hepatitis C virus.
XX
PN   JP06056891-A.
XX
PD   01-MAR-1994.
XX
PF   05-AUG-1992;   92JP-0209201.
XX
PR   05-AUG-1992;   92JP-0209201.
XX
PA   (OLYU ) OLYMPUS OPTICAL CO LTD.
XX
DR   WPI; 1994-106803/13.
XX
PT   New peptide(s) reactive with anti-hepatitis C virus antibody -
PT   for specific, early diagnosis of HCV infection
XX
PS   Disclosure; Page 9-10; 15pp; Japanese.
XX
CC   Peptide fragments of the non-structural protein (NS3) are reactive
CC   with and can detect antibodies against the NS3 domain of HCV.  The
CC   peptides can be used for diagnosis of HCV infection.  Nonspecific
CC   reacion can be inhibited and misdiagnosis of HCV infection can be
CC   decreased. See AAR51162-70.
XX
SQ   Sequence   609 AA;

  Query Match             99.0%;  Score 943;  DB 15;  Length 609;
  Best Local Similarity   98.4%;  Pred. No. 2.1e-89;
  Matches  179;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :|||||||||||'|||||||||||| ||||| |||| |||.|||||  |||||||||
Db         20 LAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 79

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              ||||||||||||||||||:|||||||||||||||||||||||||||:||||||||||.
Db         80 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 139

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              :|||||||||||||||||:|||||||||||||,|||:||||||.|||||||||':||||,
Db        140 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 199

Qy        181 RS 182
              ||
Db        200 RS 201


RESULT 6
```

```
AAP92041
ID   AAP92041 standard; protein; 1766 AA.
XX
AC   AAP92041;
XX
DT   25-MAR-2003  (updated)
DT   02-MAR-1990  (first entry)
XX
DE   Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones
DE   14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f,
DE   33f, 33g and 39c.
XX
KW   Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH)
XX
OS   Hepatitis C virus.
XX
PN   EP318216-A.
XX
PD   31-MAY-1989.
XX
PF   18-NOV-1988;   88EP-0310922.
XX
PR   18-NOV-1987;   87US-0122714.
PR   30-DEC-1987;   87US-0139886.
PR   26-FEB-1988;   88US-0161072.
PR   06-MAY-1988;   88US-0191263.
PR   26-OCT-1988;   88US-0263584.
PR   14-NOV-1988;   88US-0271450.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Houghton M, Choo QL, Kuo G;
XX
DR   WPI; 1989-159274/22.
DR   N-PSDB; AAN92097.
XX
PT   Purified hepatitis C virus
PT   - and associated nucleic acids and polypeptide(s)
XX
PS   Claim 13; Figure 26-1, 26-2, 26-3, 26-4, 26-5, 26-6; 139pp; English.
XX
CC   It is the sequence encoded in the open reading frame of hepatitis C virus
CC   cDNA inserts in clones 14i,m 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36,
CC   81, 32, 33b, 25c, 14c, 8f, a33f, 33g and 39c. It is antigenic and could
CC   be used in immunoassay reagents and vaccines and to generate antibodies
CC   useful in diagnosis and passive immunotherapy for HCV infection/non-A,
CC   non-B hepatitis.
CC   (Updated on 25-MAR-2003 to correct PR field.)
CC   (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ   Sequence   1766 AA;

  Query Match             99.0%;  Score 943;  DB 10;  Length 1766;
  Best Local Similarity   98.4%;  Pred. No. 6.8e-89;
  Matches  179;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :|||||||||||'|||||||||||| ||||| |||| |||.|||||  |||||||||
Db        310 LAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 369

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              ||||||||||||||||||:|||||||||||||||||||||||||||:||||||||||.
Db        370 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 429

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db        430 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 489

Qy        181 RS 182
              ||
Db        490 RS 491
```

```
RESULT 7
AAP90158
ID   AAP90158 standard; protein; 1786 AA.
XX
AC   AAP90158;
XX
DT   25-MAR-2003  (updated)
DT   10-NOV-1989  (first entry)
XX
DE   Protein sequence of hepatitis c virus composite cDNA.
XX
KW   Hepatitis C virus; vaccine.
XX
OS   Pan troglodytes.
XX
PN   GB2212511-A.
XX
PD   26-JUL-1989.
XX
PF   18-NOV-1988;   88GB-0027024.
XX
PR   18-NOV-1987;   87US-0122714.
PR   30-DEC-1987;   87US-0139886.
PR   26-FEB-1988;   88US-0161072.
PR   26-OCT-1988;   88US-0263584.
XX
PA   (CHIR ) CHIRON CORPORATION.
XX
PI   Houghton M,  Choo QL,  Kuo G;
XX
DR   WPI; 1989-215054/30.
DR   N-PSDB; AAN90327.
XX
PT   Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT   polypeptide(s) and antibodies for diagnosis, prevention and treatment
PT   of infection.
XX
PS   Disclosure; fig 26; 30pp; English.
XX
CC   The sequence is encoded by the composite cDNA of AAN90327.  These
CC   antigens react with antibodies in patients with non-A non-B hepatitis
CC   (NANBH). They can be used to diagnose HCV-induced NANBH, to raise
CC   antibodies for immunoassay or treatment, or to produce vaccines.
CC   (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ   Sequence   1786 AA;

  Query Match             99.0%;  Score 943;  DB 10;  Length 1786;
  Best Local Similarity   98.4%;  Pred. No. 8.9e-89;
  Matches  179;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 369

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 429

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        430 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 489

Qy        181 RS 182
              ||
Db        490 RS 491


RESULT 8
AAP90164
ID   AAP90164 standard; protein; 2261 AA.
XX
AC   AAP90164;
XX
DT   25-MAR-2003  (updated)
DT   01-NOV-1989  (first entry)
XX
DE   Peptide encoded by composite hepatitis C virus cDNA.
XX
KW   Hepatitis C virus; clone 121; clone 15e; probe; vaccine.
XX
OS   Pan troglodytes.
XX
PN   GB2212511-A.
XX
PD   26-JUL-1989.
XX
PF   18-NOV-1988;   88GB-0027024.
XX
PR   18-NOV-1987;   87US-0122714.
PR   30-DEC-1987;   87US-0139886.
PR   26-FEB-1988;   88US-0161072.
PR   26-OCT-1988;   88US-0263584.
XX
PA   (CHIR ) CHIRON CORPORATION.
XX
PI   Houghton M,  Choo QL,  Kuo G;
XX
DR   WPI; 1989-215054/30.
DR   N-PSDB; AAN90331.
XX
PT   Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT   polypeptide(s) and antibodies for diagnosis, prevention and
PT   treatment of infection.
XX
PS   Disclosure; fig 32; 235pp; English.
XX
CC   The sequence is the peptide encoded by the composite hepatitis C
CC   virus (HCV) cDNA of AAN90331.  The polypeptides are used to diagnose
CC   HCV-induced NANBH, to raise antibodies for immunoassay or treatment,
CC   or to produce vaccines.
CC   (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ   Sequence   2261 AA;

  Query Match             99.0%;  Score 943;  DB 10;  Length 2261;
  Best Local Similarity   98.4%;  Pred. No. 1.2e-88;
  Matches  179;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 460

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 520

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 580

Qy        181 RS 182
              ||
Db        581 RS 582


RESULT 9
AAP92047
ID   AAP92047 standard; protein; 2301 AA.
XX
AC   AAP92047;
XX
DT   25-MAR-2003  (updated)
DT   02-MAR-1990  (first entry)
XX
DE   Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones
```

```
DE   12f through 15e.
XX
KW   Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH).
XX
OS   Hepatitis C virus.
XX
PN   EP318216-A.
XX
PD   31-MAY-1989.
XX
PF   18-NOV-1988;   88EP-0310922.
XX
PR   18-NOV-1987;   87US-0122714.
PR   30-DEC-1987;   87US-0139886.
PR   26-FEB-1988;   88US-0161072.
PR   06-MAY-1988;   88US-0191263.
PR   26-OCT-1988;   88US-0263584.
PR   14-NOV-1988;   88US-0271450.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Houghton M,  Choo QL,  Kuo G;
XX
DR   WPI; 1989-159274/22.
DR   N-PSDB; AAN92103.
XX
PT   Purified hepatitis C virus
PT   - and associated nucleic acids and polypeptide(s)
XX
PS   Claim 13; Figure 32-1 - 32-7; 139 pp; English.
XX
CC   It is the sequence encoded in the open reading frame of hepatitis C virus
CC   (HCV) cDNA inserts in clones 12f through 15e. It is antigenic and could
CC   be used in immunoassay reagents and vaccines and to generate antibodies
CC   useful in diagnosis and passive immunotherapy for HCV infection/non-A,
CC   non-B hepatitis.
CC   (Updated on 25-MAR-2003 to correct PR field.)
CC   (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ   Sequence   2301 AA;

  Query Match             99.0%;  Score 943;  DB 10;  Length 2301;
  Best Local Similarity   98.4%;  Pred. No. 1.2e-88;
  Matches  179;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 LAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 460

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              ||||||||||||||||||||:|:|||||||||||||||||||||||||||||||||||||
Db        461 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 520

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        521 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 580

Qy        181 RS 182
              ||
Db        581 RS 582


RESULT 10
AAP92050
ID   AAP92050 standard; protein; 2436 AA.
XX
AC   AAP92050;
XX
DT   25-MAR-2003  (updated)
DT   02-MAR-1990  (first entry)
XX
DE   Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones
DE   K9-1 through 15e.
XX
KW   Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH)
XX
OS   Hepatitis C virus.
XX
PN   EP318216-A.
XX
PD   31-MAY-1989.
XX
PF   18-NOV-1988;   88EP-0310922.
XX
PR   18-NOV-1987;   87US-0122714.
PR   30-DEC-1987;   87US-0139886.
PR   26-FEB-1988;   88US-0161072.
PR   06-MAY-1988;   88US-0191263.
PR   26-OCT-1988;   88US-0263584.
PR   14-NOV-1988;   88US-0271450.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Houghton M,  Choo QL,  Kuo G;
XX
DR   WPI; 1989-159274/22.
DR   N-PSDB; AAN92106.
XX
PT   Purified hepatitis C virus
PT   - and associated nucleic acids and polypeptide(s)
XX
PS   Claim 13; Figure 47-1 - 47-8; 139 pp; English.
XX
CC   It is the sequence encoded in the open reading frame of hepatitis C virus
CC   (HCV) cDNA inserts in clones K9-1 through 15e. It is antigenic and could
CC   be used in immunoassay reagents and vaccines and to generate antibodies
CC   useful in diagnosis and passive immunotherapy for HCV infection/non-A,
CC   non-B hepatitis.
CC   (Updated on 25-MAR-2003 to correct PR field.)
CC   (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ   Sequence   2436 AA;

  Query Match             99.0%;  Score 943;  DB 10;  Length 2436;
  Best Local Similarity   98.4%;  Pred. No. 1.3e-88;
  Matches  179;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        576 LAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 635

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              ||||||||||||||||||||:|:|||||||||||||||||||||||||||||||||||||
Db        636 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 695

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        696 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 755

Qy        181 RS 182
              ||
Db        756 RS 757


RESULT 11
AAP90288
ID   AAP90288 standard; protein; 2436 AA.
XX
AC   AAP90288;
XX
DT   25-MAR-2003  (updated)
DT   19-JUL-2001  (updated)
DT   01-NOV-1989  (first entry)
XX
DE   Peptide encoded by composite hepatitis C cDNA.
XX
```

```
KW   Hepatitis C virus; clone 15e; clone k9-1; probe; vaccine.
XX
OS   Pan troglodytes.
XX
PN   GB2212511-A.
XX
PD   26-JUL-1989.
XX
PF   18-NOV-1988;   88GB-0027024.
XX
PR   18-NOV-1987;   87US-0122714.
PR   30-DEC-1987;   87US-0139886.
PR   26-FEB-1988;   88US-0161072.
PR   26-OCT-1988;   88US-0263584.
XX
PA   (CHIR ) CHIRON CORPORATION.
XX
PI   Houghton M,  Choo QL,  Kuo G;
XX
DR   WPI; 1989-215054/30.
DR   N-PSDB; AAN90336.
XX
PT   Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT   polypeptide(s) and antibodies for diagnosis, prevention and
PT   treatment of infection.
XX
PS   Disclosure; fig 47-1 to 47-8; 235pp; English.
XX
CC   The sequence is the peptide encoded by the composite hepatitis C
CC   virus (HCV) cDNA of AAN90336.  The polypeptides are used to
CC   diagnose HCV-induced NANBH, to raise antibodies for
CC   immunoassay or treatment, or to produce vaccines.
CC   (N.B. This record was resubmitted to correct errors in the sequence.)
CC   (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ   Sequence   2436 AA;

  Query Match             99.0%;  Score 943;  DB 10;  Length 2436;
  Best Local Similarity   98.4%;  Pred. No. 1.3e-88;
  Matches  179;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        576 LAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 635

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              |||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db        636 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 695

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        696 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 755

Qy        181 RS 182
              ||
Db        756 RS 757


RESULT 12
AAB18540
ID   AAB18540 standard; Protein; 2772 AA.
XX
AC   AAB18540;
XX
DT   15-JAN-2001  (first entry)
XX
DE   Protein encoded by a cDNA compiled Hepatitis C virus cDNA clones.
XX
KW   Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
KW   viral infectivity; viral replication.
XX
OS   Hepatitis C virus.
XX
PN   EP1034785-A2.
XX
PD   13-SEP-2000.
XX
PF   16-MAR-1990; 2000EP-0109602.
XX
PR   17-MAR-1989;   89US-0325338.
PR   20-APR-1989;   89US-0341334.
PR   18-MAY-1989;   89US-0355002.
PR   16-MAR-1990;   90EP-0302866.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Houghton M,  Choo Q,  Kuo G;
XX
DR   WPI; 2000-566891/53.
DR   N-PSDB; AAA75296.
XX
PT   Novel composition comprising a hepatitis C virus antisense
PT   polynucleotide which is complementary to or corresponds to a sense
PT   strand of the virus genome, and selectively hybridises to it -
XX
PS   Example; Fig 16; 75pp; English.
XX
CC   The specification describes a pharmaceutical composition which
CC   comprises a hepatitis C virus (HCV) antisense polynucleotide. The
CC   HCV is characterized by a positive stranded RNA genome which has
CC   40% homology at the polypeptide level to a HCV polyprotein. The
CC   antisense polynucleotide binds to cellular polynucleotides which
CC   enhance and/or are required for viral infectivity, replicative
CC   ability or chronicity. The antisense polynucleotides may also be
CC   designed to bind with high specificity, to be of increased stability,
CC   to be stable and to have low toxicity. The composition also comprises
CC   an agent which causes viral RNA to be inactive. The composition
CC   is used for preventing HCV replication in a system. The present
CC   sequence is encoded by a novel HCV cDNA sequence, which is used in the
CC   course of the invention.
XX
SQ   Sequence   2772 AA;

  Query Match             99.0%;  Score 943;  DB 21;  Length 2772;
  Best Local Similarity   98.4%;  Pred. No. 1.6e-88;
  Matches  179;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 LAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 971

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              |||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db        972 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 1031

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db       1032 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 1091

Qy        181 RS 182
              ||
Db       1092 RS 1093


RESULT 13
AAR34009
ID   AAR34009 standard; Protein; 2816 AA.
XX
AC   AAR34009;
XX
DT   25-MAR-2003  (updated)
DT   26-JUL-1993  (first entry)
XX
DE   HCV-1 polyprotein.
XX
KW   Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
```

```
KW   HCV; asymptomatic; chronically infected; epitope; viral isolate;
KW   domain; immunological; cross-reactive; envelope protein; vaccine;
KW   gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
XX
OS   Hepatitis C virus.
XX
PN   WO9306126-A1.
XX
PD   01-APR-1993.
XX
PF   11-SEP-1992;   92WO-US07683.
XX
PR   13-SEP-1991;   91US-0759575.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Houghton M,  Weiner AJ;
XX
DR   WPI; 1993-117468/14.
XX
PT   Immuno-reactive hepatitis C virus polypeptide compsns. - contg.
PT   at least 2 sequences from the first variable domain of distinct
PT   HCV isolates
XX
PS   Disclosure; Fig 9; 106pp; English.
XX
CC   This sequence represents the entire hepatitis C virus polyprotein.
CC   HCV is a member of the flavivirus family and appears to encode a basic
CC   polypeptide domain ("C") at the N-terminal of the viral polyprotein,
CC   followed by two glycoprotein domains ("E1", "E2/NS1"), upstream of the
CC   nonstructural genes NS2 through NS5.  See also AAQ39134-48, AAR33982-
CC   4008 and AAR38088-89.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   2816 AA;

  Query Match             99.0%;  Score 943;  DB 14;  Length 2816;
  Best Local Similarity   98.4%;  Pred. No. 1.6e-88;
  Matches  179;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :|:|||||:|||||:|||||||||||||||||||||||||||||||||||||||||||||
Db       1026 LAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              |||||:|||||:|||||:|||||||||||||||||||||||||||||||||||||||||||
Db       1086 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              ||||||||||||||||||| |||| : ||| |||||:| |||:| |||||:||:. |||
Db       1146 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 1205

Qy        181 RS 182
              ||
Db       1206 RS 1207


RESULT 14
AAR70230
ID   AAR70230 standard; protein; 2894 AA.
XX
AC   AAR70230;
XX
DT   25-MAR-2003  (updated)
DT   07-NOV-1995  (first entry)
XX
DE   Composite hepatitis C virus (HC-J1/CDC/CHI).
XX
KW   Composite hepatitis C virus; HC-J1/CDC/CHI; HCV; non-A non-B;
KW   synthetic antigens; blood screening.
XX
OS   Hepatitis C virus.
XX
PN   EP644202-A1.
XX
PD   22-MAR-1995.
XX
PF   14-DEC-1990;   94EP-0108611.
XX
PR   14-DEC-1990;   90EP-0124241.
PR   14-DEC-1990;   90EP-0108611.
XX
PA   (INNO-) INNOGENETICS NV.
XX
PI   Deleys RJ,  Maertens G,  Pollet D,  Van Heuverswyn H;
XX
DR   WPI; 1995-116946/16.
XX
PT   Synthetic antigens for the detection of hepatitis C virus
PT   antibodies - comprise portions of the HCV peptide sequence, for
PT   use in screening blood and blood products
XX
PS   Disclosure; Fig 1; 51pp; English.
XX
CC   AAR70230 is the composite hepatitis C virus (HC-J1/CDC/CHI) protein
CC   from which the synthetic HCV antigens described in AAR70210-R70229
CC   were derived. These synthetic antigens can be used to screen blood,
CC   or blood products for the presence HCV, they can also be used in
CC   various specific assays for the detection of HCV antibodies, and
CC   antigens, or as immunogens.
CC   (Updated on 25-MAR-2003 to correct PN field.)
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence   2894 AA;

  Query Match             99.0%;  Score 943;  DB 16;  Length 2894;
  Best Local Similarity   98.4%;  Pred. No. 1.7e-88;
  Matches  179;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :|:|||||:|||||:|||||||||||||||||||||||||||||||||||||||||||||
Db       1026 LAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              |||||:|||||:|||||:|||||||||||||||||||||||||||||||||||||||||||
Db       1086 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              ||||||||||||||||||| |||| : ||| |||||:| |||:| |||||:||:. |||
Db       1146 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 1205

Qy        181 RS 182
              ||
Db       1206 RS 1207


RESULT 15
AAY14975
ID   AAY14975 standard; protein; 2955 AA.
XX
AC   AAY14975;
XX
DT   20-MAR-2003  (updated)
DT   08-NOV-1999  (first entry)
XX
DE   Amino acid sequence of HCV-1 ORF.
XX
KW   Hepatitis C virus; HCV; J1; J7; HCV-1; non-A, non-B HCV; NANBH;
KW   HCV infection; vaccine.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 441
FT                   /note= "encoded by TT"
FT   Misc-difference 461
```

```
FT                   /note= "encoded by CCCC"
XX
PN   EP939128-A2.
XX
PD   01-SEP-1999.
XX
PF   17-SEP-1990;   99EP-0101746.
XX
PR   15-SEP-1989;   89US-0408045.
PR   21-DEC-1989;   89US-0456142.
PR   17-SEP-1990;   90EP-0310149.
XX
PA   (CHIR ) CHIRON CORP.
PA   (OYAA/) OYA A.
XX
PI   Cha T,  Han J,  Houghton M,  Irvine BD,  Kolberg JA;
PI   Miyamura T,  Saito I,  Weiner AJ;
XX
DR   WPI; 1999-480843/41.
DR   N-PSDB; AAZ07656.
XX
PT   New Hepatitis C Virus isolates, useful for diagnosis of hepatitis
PT   infections and development of vaccines
XX
PS   Disclosure; Fig 12; 132pp; English.
XX
CC   The invention provides two new isolates of hepatitis C virus (HCV), J1
CC   and J7. These two isolates comprise nucleotide and amino acid sequences
CC   that are distinct from the HCV isolate HCV-1. The nucleotide sequences
CC   may be used to detect non-A, non-B HCV (NANBH) polynucleotides by
CC   hybridisation for diagnosis of NANBH infections. They may also be used to
CC   screen blood donors, donated blood and blood products for this infection.
CC   The isolates may also be used to isolate other naturally occurring
CC   variants of the virus. The polypeptides may be used as a vaccine for
CC   administration to patients to protect against infection with NANBH. The
CC   present sequence represents the amino acid sequence of HCV-1 ORF.
CC   (Updated on 20-MAR-2003 to correct PF field.)
CC   (Updated on 20-MAR-2003 to correct PR field.)
XX
SQ   Sequence   2955 AA;

  Query Match             99.0%;  Score 943;  DB 20;  Length 2955;
  Best Local Similarity   98.4%;  Pred. No. 1.7e-88;
  Matches  179;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
              :|||||||||'|'||||||| |||||:|||||||.|||| .|||||| ||||||||| |||||
Db       1026 LAPITAYAQQTRGLLGCIITSLIGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085

Qy         61 GTRTIASPKGPVIQMYTNVDKDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRR 120
              |||||||||||.|||||' |:|||.||||.|||||||||||||||'|:|||| |||||'|||||
Db       1086 GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLIPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy        121 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
              ||||:|||||:|||||||||||| |:|||| | ||||:|||||'||||||.||||| |||
Db       1146 GDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 1205

Qy        181 RS 182
              ||
Db       1206 RS 1207


Search completed: September 27, 2003, 12:18:53
Job time : 79 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 29, 2003, 18:58:12 ; Search time 32 Seconds
(without alignments)
44.642 Million cell updates/sec

Title: US-09-965-594-1_COPY_14_22
Perfect score: 44
Sequence: 1 LLGCIITSL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107572

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 99%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	97.7	3010	15	AAR53417	Blood transmissibl
2	42	95.5	273	15	AAR50072	NANBH virus antige
3	42	95.5	1051	15	AAR54066	Non-A, non-B hepat
4	42	95.5	1051	16	AAR98361	5'UTR/CORE/ENV/NS1
5	41	93.2	3010	20	AAW98022	Infectious hepati
6	41	93.2	3010	22	AAB31170	Amino acid seqienc
7	41	93.2	3010	22	AAB59174	Protein encoded by
8	40	90.9	184	21	AAY44727	Hepatitis C virus
9	40	90.9	631	20	AAW93482	HCV NS3 protein.
10	39	88.6	3010	14	AAR34580	Human hepatitis C
11	39	88.6	3011	14	AAR34468	Encoded by full-le
12	37	84.1	174	13	AAR29854	HCV NS2-NS4 peptid
13	37	84.1	213	20	AAY17888	HCV NS4A-NS3 compl
14	37	84.1	215	20	AAY17890	HCV NS4A-NS3 compl
15	37	84.1	216	20	AAY17879	HCV NS4A-NS3 compl
16	37	84.1	216	20	AAY17882	HCV NS4A-NS3 compl
17	37	84.1	216	20	AAY17883	HCV NS4A-NS3 compl
18	37	84.1	216	20	AAY17886	HCV NS4A-NS3 compl
19	37	84.1	216	20	AAY17876	HCV NS4A-NS3 compl
20	37	84.1	646	20	AAY24950	HCV NS4A-NS3 compl
21	37	84.1	665	20	AAY24941	HCV NS4A-NS3 compl
22	37	84.1	665	20	AAY24942	HCV NS4A-NS3 compl
23	37	84.1	665	20	AAY24945	HCV NS4A-NS3 compl
24	37	84.1	665	20	AAY24946	HCV NS4A-NS3 compl
25	37	84.1	667	20	AAY17891	HCV NS4A-NS3 compl
26	36	81.8	300	22	AAB87799	Mouse T2R05 amino
27	36	81.8	2894	13	AAR24440	Composite HCV HC-J
28	35	79.5	76	21	AAB34473	Human secreted pro
29	35	79.5	354	21	AAG19987	Arabidopsis thalia
30	35	79.5	354	21	AAG40288	Arabidopsis thalia
31	35	79.5	473	22	AAM23525	Rabbit EST encoded
32	35	79.5	473	23	ABP69510	Human polypeptide
33	35	79.5	876	23	ABP65529	Bifidobacterium lo
34	34	77.3	144	23	ABP10525	Human ORFX protein
35	34	77.3	181	22	AAU03665	Group B Streptococ
36	34	77.3	339	23	ABP27696	Streptococcus poly
37	34	77.3	340	23	ABP27695	Streptococcus poly
38	34	77.3	345	21	AAG29992	Arabidopsis thalia
39	34	77.3	356	22	ABG21362	Novel human diagno
40	34	77.3	946	22	AAG70686	S cerevisiae apopt
41	34	77.3	1085	22	ABG27947	Novel human diagno
42	34	77.3	1087	22	ABG25216	Novel human diagno
43	33	75.0	82	23	ABG64761	Human albumin fusi
44	33	75.0	82	23	ABG64762	Human albumin fusi
45	33	75.0	82	23	ABG64945	Human albumin fusi

ALIGNMENTS

RESULT 1
AAR53417
ID AAR53417 standard; Protein; 3010 AA.
XX
AC AAR53417;
XX
DT 17-JAN-1995 (first entry)
XX
DE Blood transmissible NANBHV protein.
XX
KW Polymerase chain reaction; PCR; amplify; primer; non-A, non-B hepatitis;
KW NANBH; virus; blood transmissible; detection; hepatitis virus; RT-PCR;
KW C100 antibody; HCV RNA; NS5 region.
XX
OS Non-A, non-B hepatitis virus.
XX
FH Key Location/Qualifiers
FT Misc-difference 222
FT /label= His, Arg
FT Misc-difference 226
FT /label= Cys, Arg
FT Misc-difference 246
FT /label= Leu, Phe
FT Misc-difference 263
FT /label= Asp, Asn
FT Misc-difference 291
FT /label= Phe, Ser
FT Misc-difference 311
FT /label= Gly, Asp
FT Misc-difference 398
FT /label= Ser, Arg, Gly
FT Misc-difference 400

```
FT                   /label= Thr, Ala
FT   Misc-difference 405
FT                   /label= Gln, Pro, Leu
FT   Misc-difference 410
FT                   /label= Lys, Arg
FT   Misc-difference 418
FT                   /label= Gly, Asp
FT   Misc-difference 430
FT                   /label= Asn, Asp
FT   Misc-difference 436
FT                   /label= Phe, Leu
FT   Misc-difference 478
FT                   /label= Arg, Lys
FT   Misc-difference 759
FT                   /label= Leu, Val
FT   Misc-difference 1017
FT                   /label= Ser, Asn
FT   Misc-difference 1036
FT                   /label= Thr, Ala
FT   Misc-difference 1056
FT                   /label= Glu, Asp
FT   Misc-difference 1201
FT                   /label= Met, Thr
FT   Misc-difference 1205
FT                   /label= Met, Ile
FT   Misc-difference 1255
FT                   /label= Asn, Tyr
FT   Misc-difference 1263
FT                   /label= Gly, Asp
FT   Misc-difference 1455
FT                   /label= Asn, Asp
FT   Misc-difference 1828
FT                   /label= Ala, Thr
FT   Misc-difference 1895
FT                   /label= Gly, Arg
FT   Misc-difference 1896
FT                   /label= Gly, Ile
FT   Misc-difference 2143
FT                   /label= Glu, Val
FT   Misc-difference 2144
FT                   /label= Asp, Gln
FT   Misc-difference 2462
FT                   /label= Cys, Arg
FT   Misc-difference 2486
FT                   /label= Val, Met
FT   Misc-difference 2488
FT                   /label= Lys, Gln
FT   Misc-difference 2844
FT                   /label= Leu, Met
FT   Misc-difference 2862
FT                   /label= Leu, Gln
FT   Misc-difference 2917
FT                   /label= Arg, Leu
FT   Misc-difference 2968
FT                   /label= Ser, Gly
FT   Misc-difference 2989
FT                   /label= Cys, Arg
FT   Misc-difference 2990
FT                   /label= Tyr, Cys
XX
PN   JP05105690-A.
XX
PD   19-APR-1994.
XX
PF   10-MAR-1992;   92JP-0051885.
XX
PR   10-MAR-1992;   92JP-0051885.
XX
PA   (KAEN/) KAENNO K.
XX
DR   WPI; 1994-163130/20.
DR   N-PSDB; AAQ63499.
XX
PT   Blood-transmissible non-A non-B hepatitis virus DNA - used for
PT   detection of hepatitis virus
XX
PS   Claim 1; Page 8-20; 22pp; Japanese.
XX
CC   This sequence is encoded by the genome of a blood transmissible non-A,
CC   non-B hepatitis (NANBH) virus. The cDNA sequence was isolated using the
CC   primers given in AAQ63500-35.  The amplified fragments are used in the
CC   detection of hepatitis virus.  The target DNA was isolated from serum
CC   of chronically infected NANBH patients who were C100 antibody-positive
CC   and HCV RNA (NS5 region) positive.  Reverse transcription-PCR and PCR
CC   were performed on cDNA and the total human NANBH DNA was constructed
CC   from 23 clones.
XX
SQ   Sequence   3010 AA;

  Query Match             97.7%;  Score 43;  DB 15;  Length 3010;
  Best Local Similarity   88.9%;  Pred. No. 65;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              |||||:|||
Db       1039 LLGCIVTSL 1047


RESULT 2
AAR50072
ID   AAR50072 standard; Protein; 273 AA.
XX
AC   AAR50072;
XX
DT   24-NOV-1994  (first entry)
XX
DE   NANBH virus antigenic fragment #4.
XX
KW   Antigen; structural; non-structural; non A non B hepatitis virus;
KW   NANBHV; NANBH; patient; plasma; diagnosis; detection; carrier; ss.
XX
OS   Non A Non B hepatitis virus.
XX
PN   JP06070778-A.
XX
PD   15-MAR-1994.
XX
PF   31-JUN-1993;   93JP-0156087.
XX
PR   10-JUL-1992;   92JP-0207391.
XX
PA   (KOKU-) KOKUSAI SHIYAKU KK.
PA   (SANW ) SANWA KAGAKU KENKYUSHO CO.
PA   (TOFU ) TONEN CORP.
PA   (TOKR-) ZH TOKYOTO RINSHO IGAKU SOGO KENKYUSHO.
XX
DR   WPI; 1994-128677/16.
DR   N-PSDB; AAQ58817.
XX
PT   Nucleic acid fragment coding non-A non-B hepatitis virus antigen
PT   - useful in diagnosis of NANB patient and detection of virus
PT   carrier
XX
PS   Claim 8; Page 18-19; 37pp; Japanese.
XX
CC   The sequences given in AAR50068 and AAR50070-82 represent antigens of
CC   structural and non-structural regions of non A non B hepatitis virus
CC   (NANBHV).  The cDNA encoding these sequences wase derived from the
CC   plasma of a NANBH patient by recombinant DNA techniques.  These
CC   fragments are useful for the diagnosis of NANBH patients and the
CC   detection of NANBHV carriers.
XX
SQ   Sequence   273 AA;

  Query Match             95.5%;  Score 42;  DB 15;  Length 273;
  Best Local Similarity   88.9%;  Pred. No. 9.5;
```

```
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              |:|||||:|
Db         23 LIGCIITSL 31


RESULT 3
AAR54066
ID   AAR54066 standard; Protein; 1051 AA.
XX
AC   AAR54066;
XX
DT   14-FEB-1995  (first entry)
XX
DE   Non-A, non-B hepatitis virus gene #4 product.
XX
KW   Non-A, non-B hepatitis virus; NANBHV; hepatitis C virus; HCV;
KW   core; ENV; NS1; NS2; NS3; antigen; detection.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   Region          1..191
FT                   /label= core
FT   Region          192..383
FT                   /label= ENV
FT   Region          384..810
FT                   /label= NS1
FT   Region          811..1051
FT                   /label= NS2-NS3
FT                   /note= "NS3 N-terminal"
XX
PN   JP06141870-A.
XX
PD   24-MAY-1994.
XX
PF   12-MAR-1992;   92JP-0088140.
XX
PR   12-MAR-1992;   92JP-0088140.
XX
PA   (SANW ) SANWA KAGAKU KENKYUSHO CO.
PA   (TOFU ) TONEN CORP.
PA   (TOKR-) ZH TOKYOTO RINSHO IGAKU SOGO KENKYUSHO.
XX
DR   WPI; 1994-205028/25.
DR   N-PSDB; AAQ64068.
XX
PT   DNA coding a Non-A, non-B hepatitis virus antigen - useful for
PT   detecting HCV within serum
XX
PS   Claim 1-5; Page 11-15; 22pp; Japanese.
XX
CC   Hepatitis C virus #4 and #6 genes were isolated (AAQ64068-69).
CC   Both genes contain the core, ENV, NS1, NS2 and NS3 regions.
CC   A core region fragment is given in AAQ64067.
XX
SQ   Sequence   1051 AA;

  Query Match             95.5%;  Score 42;  DB 15;  Length 1051;
  Best Local Similarity   88.9%;  Pred. No. 35;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              |:|||.|||
Db       1039 LIGCIITSL 1047


RESULT 4
AAR98361
ID   AAR98361 standard; Protein; 1051 AA.
XX
AC   AAR98361;
XX
DT   22-AUG-1996  (first entry)
XX
DE   5'UTR/CORE/ENV/NS1/NS2/NS3 from HCV (#4).
XX
KW   Hepatitis C virus; HCV; antigen; detection; antibody.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..191
FT                   /label= Core peptide
FT   Peptide         192..383
FT                   /label= ENV1
FT   Peptide         384..810
FT                   /label= NS1/NS2
FT   Peptide         811..1051
FT                   /label= NS2 and NS3
XX
PN   JP07133291-A.
XX
PD   23-MAY-1995.
XX
PF   18-JUN-1993;   93JP-0147944.
XX
PR   18-JUN-1993;   93JP-0147944.
XX
PA   (TOFU ) TONEN CORP.
XX
DR   WPI; 1995-220780/29.
DR   N-PSDB; AAT30386.
XX
PT   Recombinant polypeptide comprising partial NS1 region of hepatitis
PT   non-A non-B viral antigen - used in a method for detecting
PT   antibodies against hepatitis non-A non-B virus.
XX
PS   Disclosure; Page 10-12; 15pp; Japanese.
XX
CC   The sequences given in AAR98361-62 rerpesent the 5'UTR/CORE/ENV/NS1/NS2/
CC   NS3 protein region derived from hepatitis C virus (HCV) isolates #4
CC   and #6 respectively.  The proteins partic. contain amino acids 384-
CC   495 of the HCV NS1 antigen.  These protein fragments may be used in
CC   the detection of antibodies against HCV.
XX
SQ   Sequence   1051 AA;

  Query Match             95.5%;  Score 42;  DB 16;  Length 1051;
  Best Local Similarity   88.9%;  Pred. No. 35;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              |:|||.|||
Db       1039 LIGCIITSL 1047


RESULT 5
AAW98022
ID   AAW98022 standard; Protein; 3010 AA.
XX
AC   AAW98022;
XX
DT   21-JUN-1999  (first entry)
XX
DE   Infectious hepatitis C virus genotype 1b strain HC-J4 protein.
XX
KW   HCV; infectious clone; infection; diagnosis; therapy; vaccine;
KW   screening; assay; antiviral; virucide.
XX
OS   Hepatitis C virus.
XX
PN   WO9904008-A2.
XX
PD   28-JAN-1999.
```

XX
PF 16-JUL-1998; 98WO-US14688.
XX
PR 27-JAN-1998; 98US-0014418.
PR 18-JUL-1997; 97US-0053092.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Bukh J, Emerson SU, Purcell RH, Yanagi M;
XX
DR WPI; 1999-132252/11.
DR N-PSDB; AAX24843.
XX
PT New isolated hepatitis C virus nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of HCV
PT infections and for developing screening assays
XX
PS Claim 2; Fig 14G-H; 126pp; English.
XX
CC This protein is encoded by the infectious hepatitis C virus (HCV)
CC genotype 1b strain HC-J4 genome (see AAX24833). HC-J4 was obtained
CC from acute phase plasma of a chimpanzee infected with serum
CC containing HC-J4/91. The infectious nucleic acid sequence can be
CC used to produce chimeric genomes (see AAX24833) consisting of the
CC open reading frames of infectious nucleic acid sequences of other
CC genotypes (including genotypes 1-6) and subtypes (such as 1b, 2a,
CC 2b, 2c, 3a, 4a-f, 5a and 6a) of HCV. The invention also relates to
CC the introduction of mutations or deletions into infectious nucleic
CC acid sequences in order to produce an attenuated HCV virus suitable
CC for vaccine development. Infectious nucleic acid sequences can also
CC be used to produce attenuated virus via passage in vitro or in vivo
CC of the viruses produced by transfection of a host cell with the
CC infectious nucleic acid sequence. Vaccines comprising one or more
CC polypeptides made from the infectious nucleic acid sequence are
CC used to immunise mammals, especially humans, against hepatitis C.
CC The nucleic acid sequences can also be used to induce protective
CC immunity against the virus. The nucleic acid sequences or their
CC encoded proteases (e.g. NS3 protease) can additionally be used to
CC develop screening assays to identify antiviral agents for HCV.
XX
SQ Sequence 3010 AA;

Query Match 93.2%; Score 41; DB 20; Length 3010;
Best Local Similarity 88.9%; Pred. No. 1.5e+02;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLGCIITSL 9
:|||||||||
Db 1039 VLGCIITSL 1047

RESULT 6
AAB31170
ID AAB31170 standard; Protein; 3010 AA.
XX
AC AAB31170;
XX
DT 02-APR-2001 (first entry)
XX
DE Amino acid sequence of a hepatitis C virus (HCV) clone genotype 1b.
XX
KW Chimeric virus; bovine viral diarrhoea virus; BVDV; hepatitis C virus;
KW HCV; vaccine; viral inhibitor; antiviral.
XX
OS Hepatitis C virus.
XX
PN WO200075352-A2.
XX
PD 14-DEC-2000.
XX
PF 02-JUN-2000; 2000WO-US15527.
XX
PR 04-JUN-1999; 99US-0137817.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Nam J, Bukh J, Emerson SU, Purcell RH;
XX
DR WPI; 2001-071081/08.
DR N-PSDB; AAC86939.
XX
PT New nucleic acid comprising a chimeric bovine viral diarrhoea virus
PT genome in which the (non-)structural region has been replaced by
PT hepatitis C virus (HCV) genome useful for treating or preventing HCV
PT signs and symptoms -
XX
PS Disclosure; Fig 4G-H; 97pp; English.
XX
CC The specification describes a nucleic acid comprising a chimeric virus
CC genome, specifically bovine viral diarrhoea virus (BVDV) genome in which
CC the (non-)structural region has been replaced by the (non-)structural
CC region of a hepatitis C virus (HCV) genome. The nucleic acids comprising
CC the chimeric virus and the chimeric virus are useful for identifying
CC cell lines capable of supporting the replication of these chimeric
CC viruses, in screening for neutralizing antibodies to HCV of different
CC genotypes, in the production of HCV-BVDV virions, for the development
CC of inactivated or attenuated vaccines to prevent HCV-BVDV in a mammal,
CC in studying the molecular properties of HCV indirectly in vitro, and in
CC identifying inhibitors of viral enzyme activity which would be useful
CC as antiviral agents. Formulations or compositions comprising the
CC chimeric virions may be used to treat or prevent the signs and symptoms
CC of HCV. The present sequence is encoded by a HCV clone, which is used
CC to construct chimeric nucleic acids of the invention.
XX
SQ Sequence 3010 AA;

Query Match 93.2%; Score 41; DB 22; Length 3010;
Best Local Similarity 88.9%; Pred. No. 1.5e+02;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLGCIITSL 9
:|||||||||
Db 1039 VLGCIITSL 1047

RESULT 7
AAB59174
ID AAB59174 standard; Protein; 3010 AA.
XX
AC AAB59174;
XX
DT 21-MAR-2001 (first entry)
XX
DE Protein encoded by infectious Hepatitis C virus 1b genotype.
XX
KW GBV-B; hepatitis C virus; HCV; vaccine.
XX
OS Hepatitis C virus.
XX
PN WO200075337-A1.
XX
PD 14-DEC-2000.
XX
PF 02-JUN-2000; 2000WO-US15293.
XX
PR 04-JUN-1999; 99US-0137694.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Bukh J, Yanagi M, Emerson SU, Purcell RH;
XX
DR WPI; 2001-091214/10.
XX
PT New infectious nucleic acids of the GB virus-B clone, useful for
PT indirectly studying the molecular properties of hepatitis C virus (HCV)
PT and in developing vaccines and therapeutics for HCV -

```
XX
PS   Disclosure; Fig 7; 96pp; English.
XX
CC   The present invention relates to GB virus-B. The nucleic acid molecules
CC   of the invention are useful for indirectly studying the molecular
CC   properties of hepatitis C virus (HCV). The infectious nucleic acid
CC   sequence of the GB virus-B clone and the HCV/GBV-B chimeras may be used
CC   in the development of vaccines and therapeutics for HCV.
XX
SQ   Sequence   3010 AA;

  Query Match             93.2%;  Score 41;  DB 22;  Length 3010;
  Best Local Similarity   88.9%;  Pred. No. 1.5e-02;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              :|||||.||
Db       1039 VLGCIITSL 1047


RESULT 8
AAY44727
ID   AAY44727 standard; Protein; 184 AA.
XX
AC   AAY44727;
XX
DT   04-MAY-2000  (first entry)
XX
DE   Hepatitis C virus NS3 catalytic domain of 1B genotype.
XX
KW   NS3 catalytic domain; NS4A peptide; NS4A-NS3 fusion construct; diagnosis;
KW   serine protease; trypsin family; screening; anti-viral compound;
KW   treatment; inhibitor; therapeutic.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   Region          1..5
FT                   /note= "Residues that are deleted for the construction
FT                   of NS4A/NS3 fusion construct"
XX
PN   WO200001718-A2.
XX
PD   13-JAN-2000.
XX
PF   02-JUL-1999;   99WO-US15035.
XX
PR   02-JUL-1998;   98US-0091675.
XX
PA   (UYFL ) UNIV FLORIDA.
XX
PI   Dunn BM,  Bukhtiyarova M;
XX
DR   WPI; 2000-182103/16.
XX
PT   Novel polypeptide comprising hepatitis C virus NS4A and NS3 domains,
PT   useful for screening for compounds useful for the diagnosis and
PT   treatment of hepatitis C virus
XX
PS   Claim 3; Fig 1; 30pp; English.
XX
CC   The present protein sequence is the catalytic domain of the NS3 clone
CC   from the 1B genotype, derived from hepatitis C virus. The NS3 protein
CC   has a sequence relationship to members of the trypsin family of serine
CC   proteases. Constructs comprising an NS4A peptide sequence covalently
CC   attached to the N-terminus of NS3 catalytic domain or their fragments
CC   can be used to create fusion polypeptides. This fusion polypeptide
CC   contains the NS3 domain expressed in a stable, soluble form. This
CC   facilitates the use of the polypeptide in direct screening of potential
CC   anti-viral compounds, that are used for diagnosis and treatment of
CC   hepatitis C virus infection. It is also used to screen for inhibitors of
CC   serine protease activity. The polynucleotides are also useful to identify
CC   diagnostic or therapeutic compounds and for recombinant production of
CC   the fusion polypeptide.
XX
SQ   Sequence   184 AA;

  Query Match             90.9%;  Score 40;  DB 21;  Length 184;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LGCIITSL 9
              ||||||||
Db         15 LGCIITSL 22


RESULT 9
AAW93482
ID   AAW93482 standard; Protein; 631 AA.
XX
AC   AAW93482;
XX
DT   11-JUN-1999  (first entry)
XX
DE   HCV NS3 protein.
XX
KW   NS3; helicase domain; X-ray crystal structure; Hepatitis C virus;
KW   HCV; computer programme; binding pocket; three-dimensional.
XX
OS   Hepatitis C virus.
XX
PN   WO9908148-A1.
XX
PD   25-FEB-1999.
XX
PF   13-AUG-1998;   98WO-US16879.
XX
PR   13-AUG-1997;   97US-0055772.
XX
PA   (VERT-) VERTEX PHARM INC.
XX
PI   Caron P,  Kim JE,  Lin C,  Morgenstern K;
XX
DR   WPI; 1999-190157/16.
DR   N-PSDB; AAX23258.
XX
PT   New hepatitis C virus NS3 helicase crystals  - which provide
PT   molecular design techniques to identify, select and design agents
PT   which bind to the helicase, particularly inhibitor compounds
XX
PS   Disclosure; Page 202-205; 224pp; English.
XX
CC   This invention relates to the X-ray crystal structure of the Hepatitis C
CC   virus helicase domain. The invention describes crystallized complexes of
CC   HCV helicase and an oligonucleotide. The described method is used in a
CC   novel computer programme where the computer is programmed with the
CC   structure coordinates of the HCV helicase oligonucleotide binding pocket
CC   or the HCV helicase nucleotide triphosphate pocket wherein the computer
CC   is capable of displaying a three-dimensional representation of that
CC   binding pocket.
XX
SQ   Sequence   631 AA;

  Query Match             90.9%;  Score 40;  DB 20;  Length 631;
  Best Local Similarity   100.0%;  Pred. No. 49;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LGCIITSL 9
              ||||.|||
Db         14 LGCIITSL 21


RESULT 10
AAR34580
ID   AAR34580 standard; Protein; 3010 AA.
XX
```

```
AC   AAR34580;
XX
DT   25-MAR-2003  (updated)
DT   25-AUG-1993  (first entry)
XX
DE   Human hepatitis C virus gene encoded polypeptide.
XX
KW   HCV; detection; diagnosis; vaccine; peptide.
XX
OS   Human hepatitis C virus.
XX
PN   EP541089-A2.
XX
PD   12-MAY-1993.
XX
PF   05-NOV-1992;   92EP-0118974.
XX
PR   07-NOV-1991;   91JP-0318679.
XX
PA   (SANW ) SANWA KAGAKU KENKYUSHO CO.
XX
PI   Hayashi Y,  Jomori T,  Kurono M,  Mitani I,  Sawai K,  Suzuki E;
XX
DR   WPI; 1993-154074/19.
DR   N-PSDB; AAQ41345.
XX
PT   Single or double stranded deoxyribonucleic acid for hepatitis C virus
PT   detection - comprises 9500 nucleotide(s) and encodes human
PT   hepatitis C virus gene, for diagnosis by polymerase chain reacting
PT   serum sample RNA deriv. and electrophoresis
XX
PS   Disclosure; Page 4-20; 21pp; English.
XX
CC   The sequence is that encoded by a human hepatitis C virus (HCV) gene.
CC   A peptide comprising part of the amino acid sequence may be used as
CC   an antigen as part of a method of detection for human anti-HCV
CC   antibody or in diagnosis of human hepatitis C. It may also be used
CC   as an antigen in a method of prodn. of vaccine for human hepatitis C.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   3010 AA;

  Query Match             88.6%;  Score 39;  DB 14;  Length 3010;
  Best Local Similarity   77.8%;  Pred. No. 3.4e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              : ||||:||
Db       1039 LFGCIVTSL 1047


RESULT 11
AAR34468
ID   AAR34468 standard; Protein; 3011 AA.
XX
AC   AAR34468;
XX
DT   30-JUL-1993  (first entry)
XX
DE   Encoded by full-length Hepatitis C virus clone JK1 5.
XX
KW   HCV; non-A, non-B hepatitis virus; NANBHV; liver disease;
KW   polymerase chain reaction; diagnostic method.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 2414
FT                   /note= "not defined"
XX
PN   JP05068562-A.
XX
PD   23-MAR-1993.
XX
PF   30-MAY-1991;   91JP-0153736.
XX
PR   30-MAY-1991;   91JP-0153736.
XX
PA   (SANW ) SANWA KAGAKU KENKYUSHO CO.
XX
DR   WPI; 1993-130638/16.
DR   N-PSDB; AAQ40425.
XX
PT   DNA and cDNA of hepatitis C virus - useful as probes for
PT   diagnosing HCV infection
XX
PS   Claim 3; Page 6-18; 44pp; Japanese.
XX
CC   cDNA was prepared from HCV genomic RNA. Full-length clone JK1-5
CC   and 14 shorter clones were isolated by PCR amplification.
CC   Primer/probes derived from the sequences of these clones can be used
CC   in diagnostic assays for HCV. See also AAQ40425-Q40439.
XX
SQ   Sequence   3011 AA;

  Query Match             88.6%;  Score 39;  DB 14;  Length 3011;
  Best Local Similarity   77.8%;  Pred. No. 3.4e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              : ||||:||
Db       1039 LFGCIVTSL 1047


RESULT 12
AAR29854
ID   AAR29854 standard; Protein; 174 AA.
XX
AC   AAR29854;
XX
DT   25-MAR-2003  (updated)
DT   26-APR-1993  (first entry)
XX
DE   HCV NS2-NS4 peptide O26-1.
XX
KW   Clone; polypeptide; NS2-NS4; Hepatitis C; Virus; HCV; serum; HC;
KW   transcriptase; cDNA; primer; allele.
XX
OS   Hepatitis C virus.
XX
PN   EP518313-A2.
XX
PD   16-DEC-1992.
XX
PF   11-JUN-1992;   92EP-0109812.
XX
PR   11-JUN-1991;   91JP-0139268.
PR   12-JUL-1991;   91JP-0172794.
PR   07-OCT-1991;   91JP-0287006.
PR   16-DEC-1991;   91JP-0332329.
PR   20-APR-1992;   92JP-0099957.
XX
PA   (MITU ) MITSUBISHI KASEI CORP.
XX
PI   Hayashi N,  Honda Y,  Murakami T,  Seki M,  Takahashi K;
PI   Teranishi Y;
XX
DR   WPI; 1992-417213/51.
DR   N-PSDB; AAQ32485.
XX
PT   New hepatitis C virus gene and its encoded protein - used for
PT   diagnosing and vaccinating against hepatitis C virus infections
XX
PS   Disclosure; Page 151-52; 305pp; English.
XX
CC   The sequences given in AAR29852-70 are encodeed by various clones which
```

CC were used in the isolation of the NS2-NS4 regions of the Hepatitis C
CC Virus (HCV) gene of the invention (see also AAR29660, AAR29559-60 and
CC AAR29843-51). These RNA sequences were isolated from the serum of a
CC patient suffering from hepatitis C (HC). The isolated RNA sequences
CC were converted into cDNA using transcriptase in the presence of one
CC of the primer sequences given in AAQ32578-79. The sequences were then
CC amplified using primer pairs. The cDNA sequences isolated represent
CC different alleles of the same region of the HCV gene. Sequence
CC comparisons of these clones showed that it is possible for a patient
CC to carry more than one HCV strain at one time. See also AAQ32436.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 174 AA;

Query Match 84.1%; Score 37; DB 13; Length 174;
Best Local Similarity 88.9%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLGCIITSL 9
||| ||||
Db 56 LLGCNITSL 64

- Not Human ?

RESULT 13
AAY17889
ID AAY17889 standard; Protein; 213 AA.
XX
AC AAY17889;
XX
DT 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex SEQ ID NO:50.
XX
KW HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO9928482-A2.
XX
PD 10-JUN-1999.
XX
PF 24-NOV-1998; 98WO-US24528.
XX
PR 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX
PA (SCHE) SCHERING CORP.
XX
PI Malcolm BA, Taremi SS, Weber PC, Yao N;
XX
DR WPI; 1999-385385/32.
XX
PT New hepatitis C virus covalent complexes
XX
PS Example 1; Page 119; 211pp; English.
XX
CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence represents an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non
CC covalent protease-peptide complexes previously available.
XX
SQ Sequence 213 AA;

Query Match 84.1%; Score 37; DB 20; Length 213;
Best Local Similarity 88.9%; Pred. No. 60;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLGCIITSL 9
||| ||||
Db 47 LLGCKITSL 55

RESULT 14
AAY17890
ID AAY17890 standard; Protein; 215 AA.
XX
AC AAY17890;
XX
DT 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex SEQ ID NO:57.
XX
KW HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO9928482-A2.
XX
PD 10-JUN-1999.
XX
PF 24-NOV-1998; 98WO-US24528.
XX
PR 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX
PA (SCHE) SCHERING CORP.
XX
PI Malcolm BA, Taremi SS, Weber PC, Yao N;
XX
DR WPI; 1999-385385/32.
XX
PT New hepatitis C virus covalent complexes
XX
PS Example 1; Page 122-123; 211pp; English.
XX
CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence represents an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.
XX
SQ Sequence 215 AA;

Query Match 84.1%; Score 37; DB 20; Length 215;
Best Local Similarity 88.9%; Pred. No. 60;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLGCIITSL 9
|||| ||||
Db 47 LLGCKITSL 55

RESULT 15
AAY17879
ID AAY17879 standard; Protein; 216 AA.

XX
AC AAY17879;
XX
DI 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex SEQ ID NO:3.
XX
KW HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO9928482-A2.
XX
PD 10-JUN-1999.
XX
PF 24-NOV-1998; 98WO-US24528.
XX
PR 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX
PA (SCHE) SCHERING CORP.
XX
PI Malcolm BA, Taremi SS, Weber PC, Yao N;
XX
DR WPI; 1999-385385/32.
XX
PT New hepatitis C virus covalent complexes
XX
PS Claim 6; Page 75-76; 211pp; English.
XX
CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence represents a specifically claimed example of the above
CC complex. The covalent NS4A-NS3 complexes are useful for structural
CC determination and determination of mode of binding of HCV inhibitors by
CC NMR spectroscopy. They can also be used for detecting inhibitors of the
CC protease activity, the helicase activity and the ATPase activity of NS3.
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
CC the non-covalent protease-peptide complexes previously available.
XX
SQ Sequence 216 AA;

Query Match 84.1%; Score 37; DB 20; Length 216;
Best Local Similarity 88.9%; Pred. No. 61;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLGCIITSL 9
|||.| ||.
Db 48 LLGCIKTSL 56

Search completed: September 29, 2003, 19:07:21
Job time : 33 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 29, 2003, 19:03:22 ; Search time: 18 Seconds
(without alignments)
48.084 Million cell updates/sec

Title: US-09-965-594-1_COPY_14_22
Perfect score: 44
Sequence: 1 LLGCIITSL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 99%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	97.7	3010	1	A45573	genome polyprotein
2	40	90.9	3010	1	GNWVTW	genome polyprotein
3	39	88.6	3010	1	S18030	genome polyprotein
4	38	86.4	231	2	T16938	hypothetical prote
5	37	84.1	851	2	G71435	hypothetical prote
6	36	81.8	403	2	B69338	conserved hypothet
7	35	79.5	653	2	A49722	endoglin precursor
8	34	77.3	179	2	AD1917	hypothetical prote
9	34	77.3	338	2	H86372	protein F508.40 [i
10	34	77.3	345	2	G86372	protein F508.39 [i
11	34	77.3	428	2	PC4163	toxin-co-regulated
12	34	77.3	946	2	S48255	probable membrane
13	34	77.3	2783	2	T31431	inositol 1,4,5-tri
14	33	75.0	350	1	S00755	pleckstrin - human
15	33	75.0	397	2	D83311	conserved hypothet
16	33	75.0	429	2	A83463	flagellar biosynth
17	33	75.0	441	2	A69744	beta-lactamase hom
18	33	75.0	472	2	G71503	probable replicati
19	33	75.0	525	2	T34544	hypothetical prote
20	33	75.0	1056	2	T02930	lysine-ketoglutara
21	33	75.0	1089	1	PFHUGA	platelet-derived g
22	32	72.7	158	2	H90459	hypothetical prote
23	32	72.7	258	2	AG0459	Sec-independent pr
24	32	72.7	301	1	S12864	retinal isomerase
25	32	72.7	323	3	C89045	protein B0288.6 [i
26	32	72.7	330	2	C69648	2-keto-3-deoxygluc
27	32	72.7	348	2	C83994	aldose 1-epimerase
28	32	72.7	367	2	C82943	ferrichrome ABC tr
29	32	72.7	393	2	D65238	hypothetical 44.4
30	32	72.7	399	2	A60088	postsynaptic membr
31	32	72.7	417	2	AB2984	acetyl-CoA C-acety
32	32	72.7	417	2	B98299	probable acyl-CoA
33	32	72.7	419	2	JC4522	alpha-galactosidas
34	32	72.7	438	2	AD0481	probable gluconate
35	32	72.7	442	2	B90122	putative helicase
36	32	72.7	444	2	PD0001	protein-glutamine
37	32	72.7	501	2	T45955	hypothetical prote
38	32	72.7	519	2	T43533	hexose transport p
39	32	72.7	531	2	T40480	hecose transporter
40	32	72.7	573	2	H86313	protein F2H15.10 [
41	32	72.7	620	2	T13460	hypothetical prote
42	32	72.7	691	1	A29996	protein-glutamine
43	32	72.7	928	2	T38419	Sec76 domain prote
44	32	72.7	1064	2	T05195	saccharopine dehyd
45	32	72.7	1074	2	JC5928	semaphorin F precu

ALIGNMENTS

RESULT 1
A45573
genome polyprotein - hepatitis C virus (strain JT)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonst
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C;Accession: A45573
R;Tanaka, T.; Kato, N.; Nakagawa, M.; Gotsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijik
Virus Res. 23, 39-53, 1992
A;Title: Molecular cloning of hepatitis C virus genome from a single Japanese carri
A;Reference number: A45573; MUID:92295714; PMID:1318627
A;Accession: A45573
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-3010 <TAN>
A;Cross-references: GB:D11168; GB:D01171; NID:g221612; PIDN:BAA01943.1; PID:g22162
A;Experimental source: HCV-JT
A;Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIP:106207)
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein;
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 97.7%; Score 43; DB 1; Length 3010;
Best Local Similarity 88.9%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 LLGCIITSL 9
              |||||:|||
Db       1039 LLGCIVTSL 1047
```

RESULT 2
GNWVTW
genome polyprotein - hepatitis C virus (strain Taiwan)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonst
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A40244

R;Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992
A;Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
A;Reference number: A40244; MUID:92230206; PMID:1314449
A;Accession: A40244
A;Molecule type: genomic RNA
A;Residues: 1-3010 <CHE>
A;Cross-references: GB:M84754
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

  Query Match 90.9%; Score 40; DB 1; Length 3010;
  Best Local Similarity 88.9%; Pred. No. 40;
  Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLGCIIITSL 9
 . ||||:|
Db 1039 LFGCIITSL 1047

RESULT 3
S18030
genome polyprotein - hepatitis C virus (isolate JK1)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Variety: isolate JK1
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
C;Accession: S18030; S33570; A48332; S18029
R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A;Description: A whole genome of hepatitis C virus cDNA was isolated from a single patie
A;Reference number: S18028
A;Accession: S18030
A;Molecule type: genomic RNA
A;Residues: 1-3010 <HON>
A;Cross-references: EMBL:X61596; NID:g59476; PIDN:CAA43723.1; PID:g59479
A;Experimental source: isolate JK1 from an individual
R;Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.
Arch. Virol. 128, 163-169, 1993
A;Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
A;Reference number: A48332; MUID:93119270; PMID:8380522
A;Accession: S33570
A;Molecule type: genomic RNA
A;Residues: 1-547,'T',549-621,'V',623-624,'S',626-652,'DL',655-761,'I',763-782 <HCW>
A;Cross-references: EMBL:X61591
A;Note: this sequence is inconsistent with the nucleotide translation
A;Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320
 as Trp, and TTC for residue 771 as Ser
A;Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBIP:121748)
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrat

  Query Match 88.6%; Score 39; DB 1; Length 3010;
  Best Local Similarity 77.8%; Pred. No. 61;
  Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLGCIITSL 9
 . |||:||
Db 1039 LFGCIVTSL 1047

RESULT 4
T16938
hypothetical protein T25D10.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T16938
R;Miller, N.
submitted to the EMBL Data Library, June 1995
A;Description: The sequence of C. elegans cosmid T25D10.
A;Reference number: Z18611
A;Accession: T16938
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-231 <MIL>
A;Cross-references: EMBL:U28409; NID:g849230; PID:g849234; PIDN:AAC46592.1; CESP:T25
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:T25D10.2
A;Introns: 35/3; 56/1; 61/3; 76/1; 105/1; 111/3; 171/1; 201/1

  Query Match 86.4%; Score 38; DB 2; Length 231;
  Best Local Similarity 87.5%; Pred. No. 11;
  Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLGCIITS 8
 ||||:|||
Db 193 LLGCLITS 200

RESULT 5
G71435
hypothetical protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
A;Variety: columbia
C;Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
C;Accession: G71435
R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.;
 P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.;
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk,
Nature 391, 485-488, 1998
A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdome
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;
C.; Chalwatzis, N.
A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis
A;Reference number: A71400; MUID:98121113; PMID:9461215
A;Accession: G71435
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-851 <BEV>
A;Cross-references: GB:Z97342; NID:g2245031; PID:e327517; PID:g2245038
C;Genetics:
A;Map position: 4COP9-4C3845

  Query Match 84.1%; Score 37; DB 2; Length 851;
  Best Local Similarity 77.8%; Pred. No. 50;
  Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLGCIITSL 9

```
              ||:| :|||
Db        291 LLGCTLTSL 299


RESULT 6
B69338
conserved hypothetical protein APC706 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C;Accession: B69338
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: B69338
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-403 <KLE>
A;Cross-references: GB:AE001056; GB:AE000782; NID:g2649379; PIDN:AAB90535.1; PID:g284991

  Query Match             81.8%;  Score 36;  DB 2;  Length 403;
  Best Local Similarity   66.7%;  Pred. No. 41;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              | |:|:|:|
Db        330 LFGCIVTAL 338


RESULT 7
A49722
endoglin precursor - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 05-Nov-1999
C;Accession: A49722; S40180
R;Yamashita, H.; Ichijo, H.; Grimsby, S.; Moren, A.; ten Dijke, P.; Miyazono, K.
J. Biol. Chem. 269, 1995-2001, 1994
A;Title: Endoglin forms a heteromeric complex with the signaling receptors for transform
A;Reference number: A49722; MUID:94124550; PMID:8294451
A;Accession: A49722
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-653 <YAM>
A;Cross-references: GB:Z23142; NID:g437919; PIDN:CAA80673.1; PID:g437920
C;Keywords: homodimer; phosphoprotein; signal transduction; transmembrane protein

  Query Match             79.5%;  Score 35;  DB 2;  Length 653;
  Best Local Similarity   87.5%;  Pred. No. 94;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LGCIITSL 9
              ||| ||||
Db        379 LGCTITSL 386


RESULT 8
AD1917
hypothetical protein all0887 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AD1917
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AD1917
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-179 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB72844.1; PID:g17130232; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all0887

  Query Match             77.3%;  Score 34;  DB 2;  Length 179;
  Best Local Similarity   85.7%;  Pred. No. 49;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIIT 7
              ||||:.|
Db         48 LLGCLIT 54


RESULT 9
H86372
protein F5O8.40 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: H86372
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; A
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewa
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; K
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Mar
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Ta
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H86372
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-338 <STO>
A;Cross-references: GB:AE005172; NID:g4056465; PIDN:AAC98038.1; GSPDB:GN00141
C;Genetics:
A;Gene: F5O8.40
A;Map position: 1

  Query Match             77.3%;  Score 34;  DB 2;  Length 338;
  Best Local Similarity   85.7%;  Pred. No. 83;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 GCIITSL 9
              ||:||||
Db        136 GCLITSL 142


RESULT 10
G86372
protein F5O8.39 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: G86372
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; A
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewa
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; K
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Mar
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Ta
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: G86372
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-345 <STO>
A;Cross-references: GB:AE005172; NID:g4056466; PIDN:AAC98039.1; GSPDB:GN00141
```

```
C;Genetics:
A;Gene: F5O8.39
A;Map position: 1

  Query Match             77.3%;  Score 34;  DB 2;  Length 345;
  Best Local Similarity   85.7%;  Pred. No. 85;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 GCIITSL 9
              ||:::||
Db        146 GCLITSL 152


RESULT 11
PC4163
toxin-co-regulated protein chain 1 - Vibrio cholerae (fragment)
C;Species: Vibrio cholerae
C;Date: 10-May-1996 #sequence_revision 19-Jul-1996 #text_change 17-Nov-2000
C;Accession: PC4163; S23261
R;Ogierman, M.A.; Voss, E.; Meaney, C.; Faast, R.; Attridge, S.R.; Manning, P.A.
Gene 170, 9-16, 1996
A;Title: Comparison of the promoter proximal regions of the toxin-co-regulated tcp gene
A;Reference number: JC4719; MUID:96200848; PMID:8621096
A;Accession: PC4163
A;Molecule type: DNA
A;Residues: 1-428 <CGI>
A;Cross-references: EMBL:X64098
R;Manning, P.A.
submitted to the EMBL Data Library, January 1992
A;Reference number: S23261
A;Accession: S23261
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-401,'RRPKN' <MAN>
A;Cross-references: EMBL:X64098; NID:g48404; PIDN:CAA45452.1; PID:g48405
C;Comment: This is a methyl-accepting chemotaxis protein.
C;Genetics:
A;Gene: tcpI

  Query Match             77.3%;  Score 34;  DB 2;  Length 428;
  Best Local Similarity   87.5%;  Pred. No. 1e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LLGCIITS 8
              || | |||
Db         11 LLACIITS 18


RESULT 12
S48255
probable membrane protein YBR086c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YBR0809
C;Species: Saccharomyces cerevisiae
C;Date: 03-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 19-Apr-2002
C;Accession: S48255; S45954; S44670
R;Mannhaupt, G.; Stucka, R.; Ehnle, S.; Vetter, I.; Feldmann, H.
Yeast 10, 1363-1381, 1994
A;Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
A;Reference number: S48255; MUID:95208357; PMID:7900426
A;Accession: S48255
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-946 <MAN>
A;Cross-references: EMBL:X78993; NID:g476045; PIDN:CAA55993.1; PID:g476046
R;Feldmann, H.; Mannhaupt, G.; Schwarzlose, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45927
A;Accession: S45954
A;Molecule type: DNA
A;Residues: 1-946 <FE2>
A;Cross-references: EMBL:Z35955; NID:g536351; PID:g536352; MIPS:YBR086c
C;Genetics:
A;Gene: SGD:IST2
A;Cross-references: SGD:S0000290
A;Map position: 2R
C;Superfamily: Saccharomyces cerevisiae probable membrane protein YBR086c
C;Keywords: transmembrane protein
F;131-147/Domain: transmembrane #status predicted <TM1>
F;158-174/Domain: transmembrane #status predicted <TM2>
F;207-243/Domain: transmembrane #status predicted <TM3>
F;248-274/Domain: transmembrane #status predicted <TM4>
F;302-324/Domain: transmembrane #status predicted <TM5>
F;450-477/Domain: transmembrane #status predicted <TM6>
F;506-532/Domain: transmembrane #status predicted <TM7>
F;563-588/Domain: transmembrane #status predicted <TM8>

  Query Match             77.3%;  Score 34;  DB 2;  Length 946;
  Best Local Similarity   62.5%;  Pred. No. 1.9e+02;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITS 8
              ::||:|::
Db        520 ILGCVIIA 527


RESULT 13
T31431
inositol 1,4,5-trisphosphate receptor, localized in plasma membrane - Panulirus arg
C;Species: Panulirus argus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
C;Accession: T31431
R;Munger, S.D.; Ache, B.W.; Greenberg, R.M.
submitted to the EMBL Data Library, March 1998
A;Description: Plasma membrane localization of an olfactory inositol 1,4,5-trispho
A;Reference number: Z21030
A;Accession: T31431
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-2783 <MUN>
A;Cross-references: EMBL:AF055079; NID:g3660666; PID:g3660667; PIDN:AAC61691.1
C;Superfamily: inositol-trisphosphate receptor

  Query Match             77.3%;  Score 34;  DB 2;  Length 2783;
  Best Local Similarity   77.8%;  Pred. No. 4.8e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              || | :|||
Db       2554 LLMCIVTSL 2562


RESULT 14
S00755
pleckstrin - human
N;Alternate names: p47; platelet/leukocyte C kinase substrate (pleckstrin)
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 20-Apr-2000
C;Accession: S00755; A45762
R;Tyers, M.; Rachubinski, R.A.; Stewart, M.I.; Varrichio, A.M.; Shorr, R.G.L.; Hasl
Nature 333, 470-473, 1988
A;Title: Molecular cloning and expression of the major protein kinase C substrate o
A;Reference number: S00755; MUID:88232910; PMID:2897630
A;Accession: S00755
A;Molecule type: mRNA
A;Residues: 1-350 <TYE>
A;Cross-references: EMBL:X07743; NID:g35517; PIDN:CAA30564.1; PID:g35518
R;Tyers, M.; Haslam, R.J.; Rachubinski, R.A.; Harley, C.B.
J. Cell. Biochem. 40, 133-145, 1989
A;Title: Molecular analysis of pleckstrin: the major protein kinase C substrate of
A;Reference number: A45762; MUID:89359547; PMID:2768345
A;Accession: A45762
A;Molecule type: mRNA
A;Residues: 1-350 <TY2>
A;Cross-references: GB:X07743; NID:g35517; PIDN:CAA30564.1; PID:g35518
A;Note: 92-Arg was also found
C;Genetics:
```

A;Gene: GDB:PLEK; p47
A;Cross-references: GDB:9836816; OMIM:173570
A;Map position: 2pter-2qter
C;Superfamily: pleckstrin; pleckstrin repeat homology
C;Keywords: phosphoprotein; signal transduction
F;3-99/Domain: pleckstrin repeat homology <PLK1>
F;243-345/Domain: pleckstrin repeat homology <PLK2>

Query Match 75.0%; Score 33; DB 1; Length 350;
Best Local Similarity 55.6%; Pred. No. 1.3e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLGCIITSL 9
Db 292 LRGCVVTSV 300

RESULT 15
D83311
conserved hypothetical protein PA2662 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: D83311
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A;Reference number: A82950; MUID:20437357; PMID:10984043
A;Accession: D83311
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-397 <STO>
A;Cross-references: GB:AE004695; GB:AE004091; NID:g9948730; PIDN:AAG06050.1; GSPDB:GN00
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA2662

Query Match 75.0%; Score 33; DB 2; Length 397;
Best Local Similarity 55.6%; Pred. No. 1.5e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLGCIITSL 9
Db 221 LLGCVLVAL 229

Search completed: September 29, 2003, 19:08:56
Job time : 22 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 29, 2003, 18:58:47 ; Search time 10 Seconds
(without alignments)
42.324 Million cell updates/sec

Title: US-09-965-594-1_COPY_14_22
Perfect score: 44
Sequence: 1 LLSCIITSL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127859

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 99%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	97.7	3010	1	POLG_HCVJT	Q00269 h genome po
2	40	90.9	3010	1	POLG_HCVIW	P29846 h genome po
3	36	81.8	300	1	T2R9_MOUSE	Q9jk44 mus musculu
4	35	79.5	653	1	EGLN_PIG	P37176 sus scrofa
5	34	77.3	946	1	YBT6_YEAST	P38250 saccharomyc
6	33	75.0	318	1	T2R7_HUMAN	Q9nyw3 homo sapien
7	33	75.0	350	1	PLEK_HUMAN	P08567 homo sapien
8	33	75.0	350	1	PLEK_MOUSE	Q9jhk5 mus musculu
9	33	75.0	406	1	ARGD_PSEPK	P59319 pseudomonas
10	33	75.0	437	1	FLHF_PSEFU	Q52256 pseudomonas
11	33	75.0	441	1	YBBE_BACSU	O05213 bacillus su
12	33	75.0	795	1	SYFB_SHEON	Q8cf99 shewanella
13	33	75.0	1089	1	PGDS_HUMAN	P16234 homo sapien
14	33	75.0	1906	1	DICE_MOUSE	Q8r418 mus musculu
15	33	75.0	1912	1	DICE_HUMAN	Q9upy3 homo sapien
16	32	72.7	222	1	PSD9_RAT	Q9wtv5 rattus norv
17	32	72.7	301	1	REIS_TODPA	P23820 todarodes p
18	32	72.7	330	1	KDGT_BACSU	P50847 bacillus su
19	32	72.7	398	1	YJGN_ECOLI	P39338 escherichia
20	32	72.7	411	1	RAPS_CHICK	O42393 gallus gall
21	32	72.7	419	1	AGAL_MOUSE	P51569 mus musculu
22	32	72.7	531	1	GHT2_SCHPO	O74969 schizosacch
23	32	72.7	689	1	TGM2_CAVCU	P08587 cavia cutle
24	32	72.7	928	1	YDG1_SCHPO	Q10491 schizosacch
25	32	72.7	1074	1	SM5A_HUMAN	Q13591 homo sapien
26	32	72.7	1077	1	SM5A_MOUSE	Q62217 mus musculu
27	31	70.5	148	1	SSRP_NEIMA	Q51111 neisseria m
28	31	70.5	162	1	IL15_BOVIN	Q28028 bos taurus
29	31	70.5	180	1	SP22_MOUSE	Q9d365 mus musculu
30	31	70.5	237	1	LAPT_PASHA	P49618 pasteurella
31	31	70.5	312	1	MCA1_HUMAN	Q12904 homo sapien
32	31	70.5	312	1	T2R6_MOUSE	P59530 mus musculu
33	31	70.5	312	1	YMI7_YEAST	Q04516 saccharomyc
34	31	70.5	334	1	YC52_PSEAE	Q91492 pseudomonas
35	31	70.5	398	1	YJGN_ECO57	P58219 escherichia
36	31	70.5	411	1	RAPS_HUMAN	Q13702 homo sapien
37	31	70.5	411	1	RAPS_MOUSE	P12672 mus musculu
38	31	70.5	411	1	RAPS_TORCA	P09108 torpedo cal
39	31	70.5	417	1	PCL_RHOCA	O69140 rhodobacter
40	31	70.5	427	1	ETIR_PIG	Q29010 sus scrofa
41	31	70.5	436	1	Y326_METJA	Q57772 methanococc
42	31	70.5	439	1	IOLF_BACSU	P42417 bacillus su
43	31	70.5	495	1	POLG_DEN23	P14339 dengue viru
44	31	70.5	556	1	PTJ_STRCO	Q9kzp1 streptomyce
45	31	70.5	609	1	THI3_YEAST	Q07471 saccharomyc

ALIGNMENTS

RESULT 1
POLG_HCVJT
ID POLG_HCVJT STANDARD; PRT; 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals.";
RL Virus Res. 23:39-53(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC --

CC --
DR EMBL; D11168; BAA01943.1; -.
DR PIR; A45573; A45573.
DR PDB; 1A1Q; 25-MAR-98.
DR PDB; 1JXP; 14-JAN-98.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR InterPro; IPR001410; DEAD.

```
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE    1083   1083      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1107   1107      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1165   1165      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND     1230   1237      ATP (POTENTIAL).
FT   SITE        1316   1319      DECH BOX.
FT   CARBOHYD     196    196      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     209    209      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     234    234      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     250    250      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     305    305      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     423    423      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     430    430      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     448    448      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     532    532      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     540    540      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     556    556      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     576    576      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     623    623      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     645    645      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2041   2041      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2077   2077      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2240   2240      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2529   2529      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2788   2788      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3010 AA;  326573 MW;  94A1C77435D642B5 CRC64;

  Query Match           97.7%;  Score 43;  DB 1;  Length 3010;
  Best Local Similarity  88.9%;  Pred. No. 5.9;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              |||||:|.|
Db       1039 LLGCIVTSL 1047


RESULT 2
POLG_HCVTW
ID   POLG_HCVTW     STANDARD;      PRT;  3010 AA.
AC   P29846;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate Taiwan) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31645;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92230206; PubMed=1314449;
RA   Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
RT   "The Taiwanese hepatitis C virus genome: sequence determination and
RT   mapping the 5' termini of viral genomic and antigenomic RNA.";
RL   Virology 188:102-113(1992).
CC   -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC       HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC       NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M84754; -; NOT_ANNOTATED_CDS.
DR   PIR; A40244; GNWVTW.
DR   PDB; 1N64; 25-FEB-03.
DR   PDB; 1NS3; 08-APR-98.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
```

```
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CORE PROTEIN (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/S2 (POTENTIAL).
FT   CHAIN       730   1006       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1210   1217       ATP (POTENTIAL).
FT   SITE       1316   1319       DEAH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    233    233       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2041   2041       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2077   2077       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2240   2240       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2529   2529       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2788   2788       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3010 AA;  327047 MW;  AAD267D55CDFE215 CRC64;

  Query Match             90.9%;  Score 40;  DB 1;  Length 3010;
  Best Local Similarity   88.9%;  Pred. No. 21;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              | ||||||| 
Db       1039 LFGCIITSL 1047


RESULT 3
T2R9_MOUSE
ID   T2R9_MOUSE     STANDARD;      PRT;   300 AA.
AC   Q9JKT4;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Taste receptor type 2 member 9 (T2R9) (Taste receptor type 2 member
DE   5) (T2R5).
GN   TAS2R5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND TOPOLOGY.
RC   STRAIN=129/SvJ;
RX   MEDLINE=20222571; PubMed=10761934;
RA   Adler E., Hoon M.A., Mueller K.L., Chandrashekar J., Ryba N.J.P.,
RA   Zuker C.S.;
RT   "A novel family of mammalian taste receptors.";
RL   Cell 100:693-702(2000).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=20222572; PubMed=10761935;
RA   Chandrashekar J., Mueller K.L., Hoon M.A., Adler E., Feng L., Guo W.,
RA   Zuker C.S., Ryba N.J.;
RT   "T2Rs function as bitter taste receptors.";
RL   Cell 100:703-711(2000).
RN   [3]
RP   REVIEW.
RX   MEDLINE=22135574; PubMed=12139982;
RA   Montmayeur J.-P., Matsunami H.;
RT   "Receptors for bitter and sweet taste.";
RL   Curr. Opin. Neurobiol. 12:366-371(2002).
RN   [4]
RP   REVIEW.
RX   MEDLINE=21634924; PubMed=11696554;
RA   Margolskee R.F.;
RT   "Molecular mechanisms of bitter and sweet taste transduction.";
RL   J. Biol. Chem. 277:1-4(2002).
RN   [5]
RP   REVIEW.
RX   MEDLINE=22469025; PubMed=12581520;
RA   Zhang Y., Hoon M.A., Chandrashekar J., Mueller K.L., Cook B., Wu D.,
RA   Zuker C.S., Ryba N.J.;
RT   "Coding of sweet, bitter, and umami tastes: different receptor cells
RT   sharing similar signaling pathways.";
RL   Cell 112:293-301(2003).
CC   -!- FUNCTION: Receptor that may play a role in the perception of
CC       bitterness and is gustducin-linked. May play a role in sensing the
CC       chemical composition of the gastrointestinal content. The activity
CC       of this receptor may stimulate alpha gustducin, mediate PLC-beta-2
CC       activation and lead to the gating of TRPM5. Functions as a ligand
CC       for cycloheximide.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Expressed in subsets of taste receptor cells
CC       of the tongue and palate epithelium and exclusively in gustducin-
CC       positive cells. Expressed in gastric and duodenal tissues.
CC   -!- MISCELLANEOUS: Most taste cells may be activated by a limited
CC       number of bitter compounds; individual taste cells can
CC       discriminate among bitter stimuli.
CC   -!- SIMILARITY: Belongs to family T2R of G-protein coupled receptors.
CC   -!- CAUTION: This protein was previously referred to T2R5, but due to
CC       its high similarity towards both the human and rat T2R9 sequences
CC       it is considered to be T2R9.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF227147; AAF43920.1; -.
DR   Pfam; PF05296; TAS2R; 1.
KW   Receptor; G-protein coupled receptor; Transmembrane.
FT   DOMAIN        1      7       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      8     28       1 (POTENTIAL).
FT   DOMAIN       29     43       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     44     64       2 (POTENTIAL).
FT   DOMAIN       65     87       EXTRACELLULAR (POTENTIAL).
```

```
FT   TRANSMEM      88    108       3 (POTENTIAL).
FT   DOMAIN       109    128       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     129    149       4 (POTENTIAL).
FT   DOMAIN       150    181       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     182    202       5 (POTENTIAL).
FT   DOMAIN       203    233       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     234    254       6 (POTENTIAL).
FT   DOMAIN       255    259       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     260    280       7 (POTENTIAL).
FT   DOMAIN       281    300       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     161    161       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   300 AA;  34416 MW;  92AD449C66FB85EB CRC64;

  Query Match             81.8%;  Score 36;  DB 1;  Length 300;
  Best Local Similarity   75.0%;  Pred. No. 16;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITS 8
              ||||::||
Db        133 LLGCLLTS 140


RESULT 4
EGLN_PIG
ID   EGLN_PIG       STANDARD;      PRT;   653 AA.
AC   P37176;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Endoglin precursor.
GN   ENG.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
RX   MEDLINE=94124550; PubMed=8294451;
RA   Yamashita H., Ichijo H., Grimsby S., Moren A., ten Dijke P.,
RA   Miyazono K.;
RT   "Endoglin forms a heteromeric complex with the signaling receptors
RT   for transforming growth factor-beta.";
RL   J. Biol. Chem. 269:1995-2001(1994).
CC   -!- FUNCTION: MAJOR GLYCOPROTEIN OF VASCULAR ENDOTHELIUM. MAY PLAY A
CC       CRITICAL ROLE IN THE BINDING OF ENDOTHELIAL CELLS TO INTEGRINS
CC       AND/OR OTHER RGD RECEPTORS.
CC   -!- SUBUNIT: HOMODIMER THAT FORMS AN HETEROMERIC COMPLEX WITH THE
CC       SIGNALING RECEPTORS FOR TRANSFORMING GROWTH FACTOR-BETA: TGF-BETA
CC       RECEPTORS I AND/OR II. IT IS ABLE TO BIND TGF-BETA 1, AND 3
CC       EFFICIENTLY AND TGF-BETA 2 LESS EFFICIENTLY.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: SOME REGIONS OF SIMILARITY TO TGF-BETA RECEPTOR TYPE
CC       III (BETAGLYCAN).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z23142; CAA80673.1; -.
DR   PIR; A49722; A49722.
DR   InterPro; IPR001507; Endoglin/CD105.
DR   Pfam; PF00100; zona_pellucida; 1.
DR   SMART; SM00241; ZP; 1.
KW   Cell adhesion; Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    653       ENDOGLIN.
FT   DOMAIN       25    581       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    582    606       POTENTIAL.
FT   DOMAIN       607    653       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       335    574       SER/THR-RICH.
FT   SITE         398    400       CELL ATTACHMENT SITE (BY SIMILARITY).
FT   CARBOHYD      57     57       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     306    306       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   653 AA;  70279 MW;  7887B4A61DFE3E5A CRC64;

  Query Match             79.5%;  Score 35;  DB 1;  Length 653;
  Best Local Similarity   87.5%;  Pred. No. 46;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LGCIITSL 9
                ||| ||||
Db        379 LGCTITSL 386


RESULT 5
YBT6_YEAST
ID   YBT6_YEAST     STANDARD;      PRT;   946 AA.
AC   P36250;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Hypothetical 105.9 kDa protein in AAC3-RFC5 intergenic region.
GN   YBR086C OR YBR0809.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=95208357; PubMed=7900426;
RA   Mannhaupt G., Stucka R., Ehnle S., Vetter I., Feldmann H.;
RT   "Analysis of a 70 kb region on the right arm of yeast chromosome II.";
RL   Yeast 10:1363-1381(1994).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X78993; CAA55593.1; -.
DR   EMBL; Z35955; CAA85034.1; -.
DR   PIR; S48255; S48255.
DR   SGD; S0000290; IST2.
DR   Pfam; PF04547; DUF590; 1.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     122    142       POTENTIAL.
FT   TRANSMEM     154    174       POTENTIAL.
FT   TRANSMEM     218    238       POTENTIAL.
FT   TRANSMEM     254    274       POTENTIAL.
FT   TRANSMEM     303    323       POTENTIAL.
FT   TRANSMEM     448    468       POTENTIAL.
FT   TRANSMEM     506    526       POTENTIAL.
FT   TRANSMEM     564    584       POTENTIAL.
SQ   SEQUENCE   946 AA;  105903 MW;  F51A43A5D378B7BC CRC64;

  Query Match             77.3%;  Score 34;  DB 1;  Length 946;
  Best Local Similarity   62.5%;  Pred. No. 96;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITS 8
                :|||:||:
Db        520 ILGCVITA 527


RESULT 6
T2R7_HUMAN
```

```
ID   T2R7_HUMAN     STANDARD;      PRT;   318 AA.
AC   Q9NYW3;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Taste receptor type 2 member 7 (T2R7) (Taste receptor family B member
DE   4) (TRB4).
GN   TAS2R7.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND TOPOLOGY.
RX   MEDLINE=20222571; PubMed=10761934;
RA   Adler E., Hoon M.A., Mueller K.L., Chandrashekar J., Ryba N.J.P.,
RA   Zuker C.S.;
RT   "A novel family of mammalian taste receptors.";
RL   Cell 100:693-702(2000).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=20222572; PubMed=10761935;
RA   Chandrashekar J., Mueller K.L., Hoon M.A., Adler E., Feng L., Guo W.,
RA   Zuker C.S., Ryba N.J.;
RT   "T2Rs function as bitter taste receptors.";
RL   Cell 100:703-711(2000).
RN   [3]
RP   REVIEW.
RX   MEDLINE=22135574; PubMed=12139982;
RA   Montmayeur J.-P., Matsunami H.;
RT   "Receptors for bitter and sweet taste.";
RL   Curr. Opin. Neurobiol. 12:366-371(2002).
RN   [4]
RP   REVIEW.
RX   MEDLINE=21634924; PubMed=11696554;
RA   Margolskee R.F.;
RT   "Molecular mechanisms of bitter and sweet taste transduction.";
RL   J. Biol. Chem. 277:1-4(2002).
RN   [5]
RP   REVIEW.
RX   MEDLINE=22469025; PubMed=12581520;
RA   Zhang Y., Hoon M.A., Chandrashekar J., Mueller K.L., Cook B., Wu D.,
RA   Zuker C.S., Ryba N.J.;
RT   "Coding of sweet, bitter, and umami tastes: different receptor cells
RT   sharing similar signaling pathways.";
RL   Cell 112:293-301(2003).
CC   -!- FUNCTION: Receptor that may play a role in the perception of
CC       bitterness and is gustducin-linked. May play a role in sensing the
CC       chemical composition of the gastrointestinal content. The activity
CC       of this receptor may stimulate alpha gustducin, mediate PLC-beta-2
CC       activation and lead to the gating of TRPM5.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Expressed in subsets of taste receptor cells
CC       of the tongue and palate epithelium and exclusively in gustducin
CC       positive cells.
CC   -!- MISCELLANEOUS: Most taste cells may be activated by a limited
CC       number of bitter compounds; individual taste cells can
CC       discriminate among bitter stimuli.
CC   -!- SIMILARITY: Belongs to family T2R of G-protein coupled receptors.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF227133; AAF43906.1; -.
DR   Genew; HGNC:14913; TAS2R7.
DR   MIM; 604793; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF05296; TAS2R; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; UNKNOWN_1.
KW   Receptor; G-protein coupled receptor; Transmembrane.
FT   DOMAIN       1      9       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     10     30      1 (POTENTIAL).
FT   DOMAIN       31     55      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     56     76      2 (POTENTIAL).
FT   DOMAIN       77     94      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     95     115     3 (POTENTIAL).
FT   DOMAIN       116    128     CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     129    149     4 (POTENTIAL).
FT   DOMAIN       150    187     EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     188    208     5 (POTENTIAL).
FT   DOMAIN       209    235     CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     236    256     6 (POTENTIAL).
FT   DOMAIN       257    266     EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     267    287     7 (POTENTIAL).
FT   DOMAIN       288    318     CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     167    167     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     175    175     N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   318 AA;  36549 MW;  F187ADB2D8274B8A CRC64;

  Query Match          75.0%;  Score 33;  DB 1;  Length 318;
  Best Local Similarity   55.6%;  Pred. No. 59;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              |||||:: |:
Db        134 LLGCVVLSV 142


RESULT 7
PLEK_HUMAN
ID   PLEK_HUMAN     STANDARD;      PRT;   350 AA.
AC   P08567; Q8WV81;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Pleckstrin (Platelet p47 protein).
GN   PLEK OR P47.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88232910; PubMed=2897630;
RA   Tyers M., Rachubinski R.A., McCaw M.L., Varrichio A.M.,
RA   Shorr R.G.L., Haslam R.J., Harley C.B.;
RT   "Molecular cloning and expression of the major protein kinase C
RT   substrate of platelets.";
RL   Nature 333:470-473(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89359547; PubMed=2768345;
RA   Tyers M., Haslam R.J., Rachubinski R.A., Harley C.B.;
RT   "Molecular analysis of pleckstrin: the major protein kinase C
RT   substrate of platelets.";
RL   J. Cell. Biochem. 40:133-145(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymph;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
```

```
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   STRUCTURE BY NMR OF 1-105.
RX   MEDLINE=94268557; PubMed=8208296;
RA   Yoon H.S., Hajduk P.J., Petros A.M., Olejniczak E.T., Meadows R.P.,
RA   Fesik S.W.;
RT   "Solution structure of a pleckstrin-homology domain.";
RL   Nature 369:672-675(1994).
CC   -!- FUNCTION: MAJOR PROTEIN KINASE C SUBSTRATE OF PLATELETS. ITS
CC       EXACT FUNCTION IS NOT KNOWN.
CC   -!- SIMILARITY: Contains 1 DEP domain.
CC   -!- SIMILARITY: Contains 2 PH domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X07743; CAA30564.1; -.
DR   EMBL; BC018549; AAH18549.1; -.
DR   PIR; S00755; S00755.
DR   PDB; 1PLS; 03-JUN-95.
DR   Genew; HGNC:9070; PLEK.
DR   MIM; 173570; -.
DR   GO; GO:0005509; F:calcium ion binding activity; TAS.
DR   InterPro; IPR000591; DEP.
DR   InterPro; IPR001849; PH.
DR   Pfam; PF00610; DEP; 1.
DR   Pfam; PF00169; PH; 2.
DR   SMART; SM00049; DEP; 1.
DR   SMART; SM00233; PH; 2.
DR   PROSITE; PS50186; DEP; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 2.
KW   Phosphorylation; Repeat; 3D-structure.
FT   DOMAIN        4    101       PH 1.
FT   DOMAIN      136    221       DEP.
FT   DOMAIN      244    347       PH 2.
FT   VARIANT      92     92       W -> R.
FT                                /FTId=VAR_005524.
FT   CONFLICT     97     97       N -> K (IN REF. 3).
FT   STRAND        7     14
FT   STRAND       22     29
FT   TURN         30     31
FT   STRAND       32     36
FT   TURN         39     40
FT   STRAND       46     49
FT   STRAND       55     56
FT   STRAND       58     73
FT   TURN         74     76
FT   STRAND       77     82
FT   HELIX        86    103
FT   TURN        104    105
SQ   SEQUENCE   350 AA;  40082 MW;  2E2A128C8B526361 CRC64;

  Query Match           75.0%;  Score 33;  DB 1;  Length 350;
  Best Local Similarity   55.6%;  Pred. No. 63;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              | |:::||:
Db        292 LRGCVVTSV 300
```

```
RESULT 8
PLEK_MOUSE
ID   PLEK_MOUSE     STANDARD;      PRT;   350 AA.
AC   O9JHK5; Q9ERI9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Pleckstrin.
GN   PLEK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RX   MEDLINE=20318622; PubMed=10860665;
RA   Cmarik J.L., Hegamyer G., Gerrard B., Dean M., Colburn N.H.;
RT   "cDNA cloning and mapping of mouse pleckstrin (Plek), a gene
RT   upregulated in transformation-resistant cells.";
RL   Genomics 66:204-212(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Zhang Y., Wu G., Paige C.J.;
RT   "Involvement of pleckstrin in B cell differentiation and
RT   activation.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Ahn H.-J., Cho J.-J.;
RT   "Mouse pleckstrin 1 is induced in mast cells after IgE cross-
RT   linking.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAJOR PROTEIN KINASE C SUBSTRATE OF PLATELETS, ITS
CC       EXACT FUNCTION IS NOT KNOWN.
CC   -!- SIMILARITY: Contains 1 DEP domain.
CC   -!- SIMILARITY: Contains 2 PH domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF191829; AAF75830.1; -.
DR   EMBL; AF073294; AAF72039.1; -.
DR   EMBL; AF303745; AAG29513.1; -.
DR   HSSP; P08567; 1PLS.
DR   MGD; MGI:1860485; Plek.
DR   InterPro; IPR000591; DEP.
DR   InterPro; IPR001849; PH.
DR   Pfam; PF00610; DEP; 1.
DR   Pfam; PF00169; PH; 2.
DR   SMART; SM00049; DEP; 1.
DR   SMART; SM00233; PH; 2.
DR   PROSITE; PS50186; DEP; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 2.
KW   Phosphorylation; Repeat.
FT   DOMAIN        4    101       PH 1.
FT   DOMAIN      136    221       DEP.
FT   DOMAIN      244    347       PH 2.
FT   CONFLICT    120    120       L -> Q (IN REF. 3).
FT   CONFLICT    225    225       F -> L (IN REF. 3).
SQ   SEQUENCE   350 AA;  39900 MW;  348F3CB469B7CC53 CRC64;

  Query Match           75.0%;  Score 33;  DB 1;  Length 350;
  Best Local Similarity   55.6%;  Pred. No. 63;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
```

```
                    | |;;;||:
Db        292 LRGCVVISV 300


RESULT 9
ARGD_PSEPK
ID   ARGD_PSEPK     STANDARD;      PRT;   406 AA.
AC   P59319;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Acetylornithine aminotransferase (EC 2.6.1.11) (ACOAT).
GN   ARGD OR PP4481.
OS   Pseudomonas putida (strain KT2440).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=160488;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22423060; PubMed=12534463;
RA   Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA   Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA   Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
RA   Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
RA   Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzez A.,
RA   Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
RA   Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA   Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuemmler B.,
RA   Fraser C.M.;
RT   "Complete genome sequence and comparative analysis of the
RT   metabolically versatile Pseudomonas putida KT2440.";
RL   Environ. Microbiol. 4:799-808(2002).
CC   -!- CATALYTIC ACTIVITY: N(2)-acetyl-L-ornithine + 2-oxoglutarate = N-
CC       acetyl-L-glutamate 5-semialdehyde + L-glutamate.
CC   -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC   -!- PATHWAY: Arginine biosynthesis; fourth step.
CC   -!- MISCELLANEOUS: May also have succinyldiaminopimelate
CC       aminotransferase activity, thus carrying out the fourth step in
CC       lysine biosynthesis.
CC   -!- SIMILARITY: Belongs to class-III of pyridoxal-phosphate-dependent
CC       aminotransferases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE016790; AAN70056.1; -.
DR   TIGR; PP4481; -.
DR   HAMAP; MF_01107; -; 1.
DR   Pfam; PF00202; aminotran_3; 1.
DR   TIGRFAMs; TIGR00707; argD; 1.
DR   PROSITE; PS00600; AA_TRANSFER_CLASS_3; 1.
KW   Arginine biosynthesis; Transferase; Aminotransferase;
KW   Pyridoxal phosphate; Complete proteome.
FT   BINDING     255     255       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   406 AA;  43490 MW;  5597D2F10C3E221E CRC64;

  Query Match             75.0%;  Score 33;  DB 1;  Length 406;
  Best Local Similarity   71.4%;  Pred. No. 72;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIIT 7
              |||::|
Db        341 LLGCVLT 347


RESULT 10
FLHF_PSEPU
ID   FLHF_PSEPU     STANDARD;      PRT;   437 AA.
AC   Q52256;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Flagellar biosynthesis protein flhF (Flagella associated GTP-binding
DE   protein).
GN   FLHF.
OS   Pseudomonas putida.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PRS2000;
RX   MEDLINE=98164369; PubMed=9503621;
RA   Ditty J.L., Grimm A.C., Harwood C.S.;
RT   "Identification of a chemotaxis gene region from Pseudomonas putida.";
RL   FEMS Microbiol. Lett. 159:267-273(1998).
CC   -!- FUNCTION: NECESSARY FOR FLAGELLAR BIOSYNTHESIS. MAY BE INVOLVED
CC       IN TRANSLOCATION OF THE FLAGELLUM (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF031898; AAC08059.1; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR000897; SRP54.
DR   Pfam; PF00448; SRP54; 1.
DR   Pfam; PF02881; SRP54_N; 1.
DR   ProDom; PD000819; SRP54; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00300; SRP54; FALSE_NEG.
KW   Flagella; GTP-binding.
FT   NP_BIND     225     232       GTP (BY SIMILARITY).
FT   NP_BIND     303     307       GTP (BY SIMILARITY).
FT   NP_BIND     361     364       GTP (BY SIMILARITY).
SQ   SEQUENCE   437 AA;  47512 MW;  EBF1A6A650B9B27A CRC64;

  Query Match             75.0%;  Score 33;  DB 1;  Length 437;
  Best Local Similarity   66.7%;  Pred. No. 76;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              | |||:| |
Db        355 LAGCILTKL 363


RESULT 11
YBBE_BACSU
ID   YBBE_BACSU     STANDARD;      PRT;   441 AA.
AC   O05213; O08069; Q45578;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical protein ybbE precursor.
GN   YBBE.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=97419514; PubMed=9274029;
RA   Liu H., Haga K., Yasumoto K., Ohashi Y., Yoshikawa H., Takahashi H.;
RT   "Sequence and analysis of a 31 kb segment of the Bacillus subtilis
RT   chromosome in the area of the rrnH and rrnG operons.";
RL   Microbiology 143:2763-2767(1997).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin H.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- SIMILARITY: BELONGS TO THE UPF0214 FAMILY.
CC   -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC       frameshift in position 338.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB002150; BAA19500.1; ALT_FRAME.
DR   EMBL; AB002150; BAA19501.1; ALT_FRAME.
DR   EMBL; Z99104; CAB11943.1; -.
DR   PIR; A69744; A69744.
DR   SubtiList; BG11566; ybbE.
DR   HAMAP; MF_01034; -; 1.
DR   InterPro; IPR001466; Beta_lactamase.
DR   Pfam; PF00144; beta-lactamase; 1.
KW   Hypothetical protein; Signal; Complete proteome.
FT   SIGNAL       1     25       POTENTIAL.
FT   CHAIN       26    441       HYPOTHETICAL PROTEIN YBBE.
SQ   SEQUENCE   441 AA;  49304 MW;  A90FD2AFC9015517 CRC64;

  Query Match             75.0%;  Score 33;  DB 1;  Length 441;
  Best Local Similarity   66.7%;  Pred. No. 77;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              ||||.:  !
Db        219 LLGCIVEKL 227


RESULT 12
SYFB_SHEON
ID   SYFB_SHEON     STANDARD;      PRT;   795 AA.
AC   Q8EF99;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20)
DE   (Phenylalanine--tRNA ligase beta chain) (PheRS).
GN   PHET OR SO2086.
OS   Shewanella oneidensis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC   Alteromonadaceae; Shewanella.
OX   NCBI_TaxID=70863;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MR-1;
RX   MEDLINE=22297686; PubMed=12368813;
RA   Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA   Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA   Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA   Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA   Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA   Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA   Feldblyum T.V., Smith H.O., Venter J.C., Nealson K.H., Fraser C.M.;
RT   "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT   Shewanella oneidensis.";
RL   Nat. Biotechnol. 20:1118-1123(2002).
CC   -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
CC       diphosphate + L-phenylalanyl-tRNA(Phe).
CC   -!- COFACTOR: Binds 2 magnesium ions per tetramer (By similarity).
CC   -!- SUBUNIT: Tetramer of two alpha and two beta chains (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the phenylalanyl-tRNA synthetase beta chain
CC       family. Subfamily 1.
CC   -!- SIMILARITY: Contains 1 tRNA-binding domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE015650; AAN55133.1; -.
DR   TIGR; SO2086; -.
DR   HAMAP; MF_00283; -; 1.
DR   InterPro; IPR005146; B3_4.
DR   InterPro; IPR005147; B5.
DR   InterPro; IPR005121; Fdx-AntiCB.
DR   InterPro; IPR004532; PheT_bact.
DR   InterPro; IPR002547; tRNA_bind.
DR   Pfam; PF03483; B3_4; 1.
DR   Pfam; PF03484; B5; 1.
DR   Pfam; PF03147; FDX-ACB; 1.
DR   Pfam; PF01588; tRNA_bind; 1.
DR   TIGRFAMs; TIGR00472; pheT_bact; 1.
DR   PROSITE; PS50886; TRBD; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Metal-binding; Magnesium; RNA-binding; tRNA-binding;
KW   Complete proteome.
FT   DOMAIN       39    148       TRNA-BINDING.
FT   METAL       454    454       MAGNESIUM (BY SIMILARITY).
FT   METAL       460    460       MAGNESIUM (VIA CARBONYL OXYGEN) (BY
FT                                SIMILARITY).
FT   METAL       463    463       MAGNESIUM (BY SIMILARITY).
FT   METAL       464    464       MAGNESIUM (BY SIMILARITY).
SQ   SEQUENCE   795 AA;  86574 MW;  D73CD96D4CAE1296 CRC64;

  Query Match             75.0%;  Score 33;  DB 1;  Length 795;
  Best Local Similarity   71.4%;  Pred. No. 1.3e-02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIIT 7
```

```
              :|||:|;
Db        591 MLGCVIT 597


RESULT 13
PGDS_HUMAN
ID   PGDS_HUMAN     STANDARD;      PRT;  1089 AA.
AC   P16234;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Alpha platelet-derived growth factor receptor precursor (EC 2.7.1.112)
DE   (PDGF-R-alpha) (CD140a antigen).
GN   PDGFRA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89130149; PubMed=2536956;
RA   Matsui T., Heidaran M., Miki T., Popescu N., La Rochelle W.,
RA   Kraus M., Pierce J., Aaronson S.;
RT   "Isolation of a novel receptor cDNA establishes the existence of two
RT   PDGF receptor genes.";
RL   Science 243:800-804(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89296915; PubMed=2544881;
RA   Claesson-Welsh L., Eriksson A., Westermark B., Heldin C.H.;
RT   "cDNA cloning and expression of the human A-type platelet-derived
RT   growth factor (PDGF) receptor establishes structural similarity to
RT   the B-type PDGF receptor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:4917-4921(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=96163874; PubMed=8586421;
RA   Kawagishi J., Ku T.;
RT   "Structure, organization, and transcription units of the human
RT   alpha-platelet-derived growth factor receptor gene, PDGFRA.";
RL   Genomics 30:224-232(1995).
CC   -!- FUNCTION: THIS RECEPTOR BINDS PLATELET-DERIVED GROWTH FACTOR AND
CC       HAS A TYROSINE-PROTEIN KINASE ACTIVITY. THIS RECEPTOR CAN BIND
CC       EITHER PDGF-A OR PDGF-B.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: DIMER OF EITHER ALPHA-ALPHA, BETA-BETA OR ALPHA-BETA
CC       SUBUNITS.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CSF-1/PDGF RECEPTOR FAMILY OF TYROSINE-
CC       PROTEIN KINASES.
CC   -!- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M21574; AAA96715.1; -.
DR   EMBL; M22734; AAA60048.1; -.
DR   EMBL; D50017; BAA08742.1; -.
DR   EMBL; D50002; BAA08742.1; JOINED.
DR   EMBL; D50003; BAA08742.1; JOINED.
DR   EMBL; D50004; BAA08742.1; JOINED.
DR   EMBL; D50005; BAA08742.1; JOINED.
DR   EMBL; D50006; BAA08742.1; JOINED.
DR   EMBL; D50007; BAA08742.1; JOINED.
DR   EMBL; D50008; BAA08742.1; JOINED.
DR   EMBL; D50009; BAA08742.1; JOINED.
DR   EMBL; D50010; BAA08742.1; JOINED.
DR   EMBL; D50011; BAA08742.1; JOINED.
DR   EMBL; D50012; BAA08742.1; JOINED.
DR   EMBL; D50013; BAA08742.1; JOINED.
DR   EMBL; D50014; BAA08742.1; JOINED.
DR   EMBL; D50015; BAA08742.1; JOINED.
DR   EMBL; D50016; BAA08742.1; JOINED.
DR   PIR; A40162; PFHUGA.
DR   HSSP; P11362; 1FGK.
DR   Genew; HGNC:8803; PDGFRA.
DR   MIM; 173490; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0005018; F:platelet-derived growth factor, alpha-recep. . .; TAS.
DR   GO; GO:0008283; P:cell proliferation; TAS.
DR   GO; GO:0007166; P:cell surface receptor linked signal transdu. . .; TAS.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR001824; RTKinaseIII.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00047; ig; 3.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 2.
DR   SMART; SM00406; IGc2; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS50835; IG_LIKE; 2.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00240; RECEPTOR_TYR_KIN_III; 1.
KW   Tyrosine-protein kinase; Receptor; Transmembrane; Glycoprotein;
KW   Transferase; Phosphorylation; ATP-binding; Immunoglobulin domain;
KW   Signal; Repeat.
FT   SIGNAL        1     23
FT   CHAIN        24   1089       ALPHA PLATELET-DERIVED GROWTH FACTOR
FT                                RECEPTOR.
FT   DOMAIN        24    524       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     525    549       POTENTIAL.
FT   DOMAIN       550   1089       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        24    113       IG-LIKE C2-TYPE 1.
FT   DOMAIN       202    306       IG-LIKE C2-TYPE 3.
FT   DOMAIN       319    410       IG-LIKE C2-TYPE 4.
FT   DOMAIN       593    954       PROTEIN KINASE.
FT   DOMAIN      1041   1087       SER-RICH.
FT   NP_BIND      599    607       ATP (BY SIMILARITY).
FT   BINDING      627    627       ATP (BY SIMILARITY).
FT   ACT_SITE     818    818       BY SIMILARITY.
FT   MOD_RES      849    849       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD      42     42       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      76     76       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     103    103       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     179    179       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     353    353       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     359    359       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     458    458       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     468    468       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1089 AA;  122669 MW;  5E3FB9940ACD1BE8 CRC64;

  Query Match             75.0%;  Score 33;  DB 1;  Length 1089;
  Best Local Similarity   55.6%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              :|||::| :
Db         10 VLGCLLTGL 18


RESULT 14
DICE_MOUSE
ID   DICE_MOUSE     STANDARD;      PRT;  1906 AA.
AC   Q8R418;
DT   28-FEB-2003 (Rel. 41, Created)
```

```
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Endoribonuclease Dicer (EC 3.1.26.-) (Double-strand-specific
DE   ribonuclease MDCR-1).
GN   DICER1 OR DICER OR MDCR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21886641; PubMed=11989553;
RA   Nicholson R.H., Nicholson A.W.;
RT   "Molecular characterization of a mouse cDNA encoding Dicer, a
RT   ribonuclease III ortholog involved in RNA interference.";
RL   Mamm. Genome 13:67-73(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Czech II;
RA   Svoboda P., Anger M., Stein P., Schultz R.M.;
RT   "Mouse dicer homolog in oocyte and preimplantation embryos.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Doi N., Zenno S., Ui-Tei K., Takahasi F., Ueda R., Miyata Y.,
RA   Saigo K.;
RT   "eIF2C family protains and Dicer homologs are required for siRNA
RT   mediated RNAi in mammalian cells.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Involved in cleaving double-stranded RNA in the RNA
CC       interference (RNAi) pathway. It produces 21 to 23 bp dsRNAs
CC       (siRNAs) which target the selective destruction of homologous
CC       RNAs.
CC   -!- TISSUE SPECIFICITY: Expressed in a wide variety of tissues.
CC   -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY.
CC   -!- SIMILARITY: Contains 2 RNase III domains.
CC   -!- SIMILARITY: Contains 1 PAZ domain.
CC   -!- SIMILARITY: Contains 1 DRBM (double-stranded RNA-binding) domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF484524; AAL84636.1; -.
DR   EMBL; AF430845; AAM21495.1; ALT_INIT.
DR   EMBL; AB061470; BAC15765.1; -.
DR   MGD; MGI:2177178; Dicer1.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001159; DS_RBD.
DR   InterPro; IPR005034; DUF283.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR000999; RNase_3.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00035; dsrm; 1.
DR   Pfam; PF03368; DUF283; 1.
DR   Pfam; PF00271; Helicase_C; 1.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF00636; Ribonuclease_3; 2.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00358; DSRM; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00535; RIBOc; 2.
DR   PROSITE; PS50137; DS_RBD; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS00517; RNASE_3_1; 1.
DR   PROSITE; PS50142; RNASE_3_2; 2.
KW   Helicase; ATP-binding; RNA-binding; Hydrolase; Nuclease; Endonuclease;
KW   Repeat.
FT   NP_BIND      34     41       ATP (POTENTIAL).
FT   SITE        165    168       DECH BOX.
FT   DOMAIN      881   1032       PAZ.
FT   DOMAIN     1266   1393       RNASE III 1.
FT   DOMAIN     1650   1808       RNASE III 2.
FT   DOMAIN     1833   1898       DRBM.
FT   CONFLICT     97     97       A -> C (IN REF. 2).
FT   CONFLICT    157    157       S -> P (IN REF. 3).
FT   CONFLICT    279    279       H -> Y (IN REF. 2).
FT   CONFLICT    600    600       A -> T (IN REF. 3).
FT   CONFLICT    749    749       E -> D (IN REF. 2).
FT   CONFLICT    827    827       T -> I (IN REF. 2).
FT   CONFLICT    878    878       G -> S (IN REF. 1).
FT   CONFLICT    955    955       Y -> C (IN REF. 3).
FT   CONFLICT    983    983       I -> A (IN REF. 3).
FT   CONFLICT   1080   1080       R -> G (IN REF. 3).
FT   CONFLICT   1100   1100       T -> S (IN REF. 3).
FT   CONFLICT   1326   1326       P -> H (IN REF. 1).
FT   CONFLICT   1609   1609       A -> S (IN REF. 3).
FT   CONFLICT   1850   1850       K -> E (IN REF. 3).
SQ   SEQUENCE   1906 AA;  215766 MW;  230EA9BFC19F3091 CRC64;

  Query Match             75.0%;  Score 33;  DB 1;  Length 1906;
  Best Local Similarity   75.0%;  Pred. No. 2.6e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LLGCIITS 8
              |||| :|| 
Db       1552 LLGCYLTS 1559


RESULT 15
DICE_HUMAN
ID   DICE_HUMAN     STANDARD;      PRT;  1912 AA.
AC   Q9UPY3; O95943; Q9UQ02;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Endoribonuclease Dicer (EC 3.1.26.-) (Helicase with RNase motif)
DE   (Helicase-MOI).
GN   DICER1 OR DICER OR HERNA OR KIAA0928.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20246304; PubMed=10786632;
RA   Matsuda S., Ichigotani Y., Okuda T., Irimura T., Nakatsugawa S.,
RA   Hamaguchi M.;
RT   "Molecular cloning and characterization of a novel human gene (HERNA)
RT   which encodes a putative RNA-helicase.";
RL   Biochim. Biophys. Acta 1490:163-169(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Provost P., Dishart D., Doucet D., Hermansson A., Frendewey D.,
RA   Samuelsson B., Radmark O.;
RT   "RNA binding and processing by recombinant human Dicer.";
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99246063; PubMed=10231032;
RA   Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosawa M.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIII.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:63-70(1999).
RN   [4]
RP   REVISIONS.
RX   MEDLINE=22158633; PubMed=12168954;
```

```
RA   Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT   "Construction of expression-ready cDNA clones for KIAA genes: manual
RT   curation of 330 KIAA cDNA clones.";
RL   DNA Res. 9:99-106(2002).
RN   [5]
RP   SEQUENCE OF 1238-1912 FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=99162526; PubMed=10051563;
RA   Provost P., Samuelsson B., Radmark O.;
RT   "Interaction of 5-lipoxygenase with cellular proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:1881-1885(1999).
CC   -!- FUNCTION: Involved in cleaving double-stranded RNA in the RNA
CC       interference (RNAi) pathway. It produces 21 to 23 bp dsRNAs
CC       (siRNAs) which target the selective destruction of homologous
CC       RNAs.
CC   -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY.
CC   -!- SIMILARITY: Contains 2 RNase III domains.
CC   -!- SIMILARITY: Contains 1 PAZ domain.
CC   -!- SIMILARITY: Contains 1 DRBM (double-stranded RNA-binding) domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB028449; BAA78691.1; ALT_INIT.
DR   EMBL; AJ132261; CAB38857.2; -.
DR   EMBL; AB023145; BAA76772.2; ALT_INIT.
DR   Genew; HGNC:17098; DICER1.
DR   MIM; 606241; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001159; DS_RBD.
DR   InterPro; IPR005034; DUF283.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR000999; RNase_3.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00035; dsrm; 1.
DR   Pfam; PF03368; DUF283; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF00636; Ribonuclease_3; 2.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00358; DSRM; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00535; RIBOc; 2.
DR   PROSITE; PS50137; DS_RBD; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS00517; RNASE_3_1; 1.
DR   PROSITE; PS50142; RNASE_3_2; 2.
KW   Helicase; ATP-binding; RNA-binding; Hydrolase; Nuclease; Endonuclease;
KW   Repeat.
FT   NP_BIND      34     41       ATP (POTENTIAL).
FT   SITE        165    168       DECH BOX.
FT   DOMAIN      881   1032       PAZ.
FT   DOMAIN     1266   1393       RNASE III 1.
FT   DOMAIN     1656   1814       RNASE III 2.
FT   DOMAIN     1839   1904       DRBM.
FT   CONFLICT     65     80       VLLTKELSYQIRGDFS -> STTLLKSCLYLDIGETSA
FT                                (IN REF. 1).
FT   CONFLICT    179    179       I -> F (IN REF. 1).
FT   CONFLICT    185    185       N -> I (IN REF. 1).
FT   CONFLICT    204    204       C -> W (IN REF. 1).
FT   CONFLICT    208    208       E -> D (IN REF. 1).
FT   CONFLICT    213    213       I -> F (IN REF. 1).
FT   CONFLICT    383    384       QQ -> ES (IN REF. 1).
FT   CONFLICT    482    483       KQ -> NT (IN REF. 1).
FT   CONFLICT    599    599       D -> H (IN REF. 1).
SQ   SEQUENCE   1912 AA;  217627 MW;  996399DB4B074F21 CRC64;
```

```
  Query Match           75.0%;  Score 33;  DB 1;  Length 1912;
  Best Local Similarity  75.0%;  Pred. No. 2.6e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LLGCIITS 8
              |||| :||
Db       1556 LLGCYLTS 1563


Search completed: September 29, 2003, 19:07:47
Job time : 13 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 30, 2003, 10:34:19 ; Search time 2202 Seconds
(without alignments)
167.206 Million cell updates/sec

Title: US-09-965-594-1_COPY_14_22
Perfect score: 44
Sequence: 1 LLGCIITSL 9

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5776890

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 99%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV-xlib
-Q=/cgn2_1/USPTO_spool/US09965594/runat_29092003_164249_14072/app_query.fasta_1.199
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=99.9 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=10000 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594_@CGN_1_1_3508_@runat_29092003_164249_14072 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	97.7	540	14	AB072047	AB072047 Hepatitis
2	43	97.7	540	14	AB072048	AB072048 Hepatitis
3	43	97.7	540	14	AB072067	AB072067 Hepatitis
4	43	97.7	540	14	AB072102	AB072102 Hepatitis
5	43	97.7	540	14	AB089533	AB089533 Hepatitis
6	43	97.7	540	14	AB089539	AB089539 Hepatitis
7	43	97.7	540	14	AB089558	AB089558 Hepatitis
8	43	97.7	540	14	AB100809	AB100809 Hepatitis
9	43	97.7	540	14	AB100827	AB100827 Hepatitis
10	43	97.7	1211	14	AY070171	AY070171 Hepatitis
c 11	43	97.7	2693	14	AF187824	AF187824 Cache Val
c 12	43	97.7	2691	14	AF231116	AF231116 Cache Val
c 13	43	97.7	2691	14	AF231118	AF231118 Cache Val
c 14	43	97.7	4463	14	AF186241	AF186241 Cache Val
15	43	97.7	9359	14	AF313916	AF313916 Hepatitis
16	43	97.7	9369	14	AF165053	AF165053 Hepatitis
17	43	97.7	9379	14	AF165049	AF165049 Hepatitis
18	43	97.7	9379	14	AF165050	AF165050 Hepatitis
19	43	97.7	9434	14	HPCJTB	D11355 Hepatitis C
20	43	97.7	9436	6	E07266	E07266 Blood-sprea
21	43	97.7	9436	14	HPCJTA	D11168 Hepatitis C
22	43	97.7	9441	14	AY051292	AY051292 Hepatitis
23	43	97.7	9598	14	AB049101	AB049101 Hepatitis
24	43	97.7	217285	2	AC126317	AC126317 Rattus no
c 25	43	97.7	260760	2	AC127861	AC127861 Rattus no
26	42	95.5	543	14	AF369258	AF369258 Hepatitis
27	42	95.5	819	6	E06893	E06893 DNA encodin
28	42	95.5	3461	6	E07544	E07544 Gene fragme
29	42	95.5	3461	6	E09288	E09288 cDNA encodi
30	42	95.5	9375	14	AF207754	AF207754 Hepatitis
31	42	95.5	9379	14	AF207768	AF207768 Hepatitis
c 32	42	95.5	89352	9	AC006030	AC006030 Homo sapi
33	42	95.5	106991	2	AC016960	AC016960 Homo sapi
c 34	42	95.5	133801	9	AC004585	AC004585 Homo sapi
c 35	42	95.5	138517	9	AC019288	AC019288 Homo sapi
c 36	42	95.5	230396	2	AC103315	AC103315 Rattus no
c 37	42	95.5	231275	2	AC099421	AC099421 Rattus no
38	41	93.2	540	14	AB072056	AB072056 Hepatitis
39	41	93.2	1096	14	AB013622	AB013622 Hepatitis
40	41	93.2	8780	14	AF054253	AF054253 Hepatitis
41	41	93.2	8780	14	AF054255	AF054255 Hepatitis
42	41	93.2	8780	14	AF054259	AF054259 Hepatitis
43	41	93.2	8781	14	AF054256	AF054256 Hepatitis
44	41	93.2	9379	14	AF207773	AF207773 Hepatitis
45	41	93.2	9448	14	HPCJ483	D13558 Hepatitis C

ALIGNMENTS

RESULT 1

```
AB072047
LOCUS       AB072047                 540 bp    RNA     linear   VRL 02-APR-2002
DEFINITION  Hepatitis C virus type 1b gene for polyprotein, NS3 region, partial
            cds, isolate:H05-4.
ACCESSION   AB072047
VERSION     AB072047.1  GI:19911838
KEYWORDS    .
SOURCE      Hepatitis C virus type 1b
  ORGANISM  Hepatitis C virus type 1b
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1
  AUTHORS   Ogata,S.
  TITLE     CORRELATION BETWEEN SECONDARY STRUCTURE OF AN AMINO-TERMINAL
            PORTION OF THE NONSTRUCTURAL PROTEIN 3 OF HEPATITIS C VIRUS AND
            DEVELOPMENT OF HEPATOCELLULAR CARCINOMA
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 540)
  AUTHORS   Ogata,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2001) Satoshi Ogata, Kobe University, Graduate
            School of Medicine, Department of Microbiology; 7-5-1 Kusunoki-cho,
            Chuo-ku, Kobe, Hyogo 650-0017, Japan
            (E-mail:ogatas@med.kobe-u.ac.jp, Tel:81-78-382-5501,
            Fax:81-78-382-5519)
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Hepatitis C virus type 1b"
                     /mol_type="genomic RNA"
                     /isolate="H05-4"
                     /db_xref="taxon:31647"
     gene            1. .540
                     /gene="NS3"
     CDS             <1. .>540
                     /gene="NS3"
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="BAB88229.1"
                     /db_xref="GI:19911839"
                     /translation="APITAYAQQTRGLLGCIVTSLTGRDKNQVEGEVQVVSTATQSFL
                     ATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWPAPPGARSLTPCTCGSS
                     DLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGVFRAAVCT
                     RGVAKAVDFVPVESMETTMR"
     mat_peptide     1. .>540
                     /gene="NS3"
                     /product="nonstructural protein"
BASE COUNT       90 a    168 c    168 g    114 t
ORIGIN

Alignment Scores:
Pred. No.:                4.14          Length:          540
Score:                    43.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              97.73%        Indels:          0
DB:                       14            Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AB072047 (1-540)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||||||||||:|||||:::|||||:|||
Db         37 CTGCTTGGTTGCATTGTCACTAGCCTC 63

RESULT 2
AB072048
LOCUS       AB072048                 540 bp    RNA     linear   VRL 02-APR-2002
DEFINITION  Hepatitis C virus type 1b gene for polyprotein, NS3 region, partial
            cds, isolate:H05-5.
ACCESSION   AB072048
VERSION     AB072048.1  GI:19911840
KEYWORDS    .
SOURCE      Hepatitis C virus type 1b
  ORGANISM  Hepatitis C virus type 1b
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1
  AUTHORS   Ogata,S.
  TITLE     CORRELATION BETWEEN SECONDARY STRUCTURE OF AN AMINO-TERMINAL
            PORTION OF THE NONSTRUCTURAL PROTEIN 3 OF HEPATITIS C VIRUS AND
            DEVELOPMENT OF HEPATOCELLULAR CARCINOMA
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 540)
  AUTHORS   Ogata,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2001) Satoshi Ogata, Kobe University, Graduate
            School of Medicine, Department of Microbiology; 7-5-1 Kusunoki-cho,
            Chuo-ku, Kobe, Hyogo 650-0017, Japan
            (E-mail:ogatas@med.kobe-u.ac.jp, Tel:81-78-382-5501,
            Fax:81-78-382-5519)
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Hepatitis C virus type 1b"
                     /mol_type="genomic RNA"
                     /isolate="H05-5"
                     /db_xref="taxon:31647"
     gene            1. .540
                     /gene="NS3"
     CDS             <1. .>540
                     /gene="NS3"
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="BAB88230.1"
                     /db_xref="GI:19911841"
                     /translation="APITAYSQQTRGLLGCIVTSLTGRDKNQVEGEVQVVSTAIQSFL
                     ATCINGVCWTVYHGAGSKILAGPKGPITQMYTNVDLDLVGWQAPPGARSMTPCTCGSS
                     DLYLVTRHADVIPVRRRGDGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGVFRAAVCT
                     RGVAKAVDFIPVESMETTMR"
     mat_peptide     1. .>540
                     /gene="NS3"
                     /product="nonstructural protein"
BASE COUNT       95 a    177 c    165 g    103 t
ORIGIN

Alignment Scores:
Pred. No.:                4.14          Length:          540
Score:                    43.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              97.73%        Indels:          0
DB:                       14            Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AB072048 (1-540)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||||||||||:|||||:::|||||:|||
Db         37 CTACTCGGTTGCATCGTCACTAGCCTC 63

RESULT 3
AB072067
LOCUS       AB072067                 540 bp    RNA     linear   VRL 02-APR-2002
DEFINITION  Hepatitis C virus type 1b gene for polyprotein, NS3 region, partial
            cds, isolate:L08.
ACCESSION   AB072067
VERSION     AB072067.1  GI:19911878
KEYWORDS    .
SOURCE      Hepatitis C virus type 1b
  ORGANISM  Hepatitis C virus type 1b
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1
  AUTHORS   Ogata,S.
  TITLE     CORRELATION BETWEEN SECONDARY STRUCTURE OF AN AMINO-TERMINAL
            PORTION OF THE NONSTRUCTURAL PROTEIN 3 OF HEPATITIS C VIRUS AND
            DEVELOPMENT OF HEPATOCELLULAR CARCINOMA
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 540)
```

```
  AUTHORS   Ogata,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2001) Satoshi Ogata, Kobe University, Graduate
            School of Medicine, Department of Microbiology; 7-5-1 Kusunoki-cho,
            Chuo-ku, Kobe, Hyogo 650-0017, Japan
            (E-mail:ogatas@med.kobe-u.ac.jp, Tel:81-78-382-5501,
            Fax:81-78-382-5519)
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Hepatitis C virus type 1b"
                     /mol_type="genomic RNA"
                     /isolate="L08"
                     /db_xref="taxon:31647"
     gene            1. .540
                     /gene="NS3"
     CDS             <1. .>540
                     /gene="NS3"
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="BAB88249.1"
                     /db_xref="GI:19911879"
                     /translation="APITAYSQQTRGLLGCIVTSLTGRDKNQVEGEVQVVSTATQSFL
                     ATCINGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSS
                     DLYLVTRHADVIPVRRRGDTRGSLLSPRPISYLKGSSGGPLLCPSGHAVGIFRAAVCT
                     RGVAKAVDFIPVESMETTMR"
     mat_peptide     1. .>540
                     /gene="NS3"
                     /product="nonstructural protein"
BASE COUNT       95 a    168 c    160 g    114 t
ORIGIN

Alignment Scores:
Pred. No.:                4.14          Length:          540
Score:                    43.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              97.73%        Indels:          0
DB:                       14            Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AB072067 (1-540)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||||.| |||||| |:::|:|||| :
Db         37 CTACTTGGCTGCATTGTCACCAGCCTC 63

RESULT 4
AB072102
LOCUS       AB072102                 540 bp    RNA     linear   VRL 02-APR-2002
DEFINITION  Hepatitis C virus type 1b gene for polyprotein, NS3 region, partial
            cds, isolate:82.
ACCESSION   AB072102
VERSION     AB072102.1  GI:19911948
KEYWORDS    .
SOURCE      Hepatitis C virus type 1b
  ORGANISM  Hepatitis C virus type 1b
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1
  AUTHORS   Ogata,S.
  TITLE     CORRELATION BETWEEN SECONDARY STRUCTURE OF AN AMINO-TERMINAL
            PORTION OF THE NONSTRUCTURAL PROTEIN 3 OF HEPATITIS C VIRUS AND
            DEVELOPMENT OF HEPATOCELLULAR CARCINOMA
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 540)
  AUTHORS   Ogata,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2001) Satoshi Ogata, Kobe University, Graduate
            School of Medicine, Department of Microbiology; 7-5-1 Kusunoki-cho,
            Chuo-ku, Kobe, Hyogo 650-0017, Japan
            (E-mail:ogatas@med.kobe-u.ac.jp, Tel:81-78-382-5501,
            Fax:81-78-382-5519)
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Hepatitis C virus type 1b"
                     /mol_type="genomic RNA"
                     /isolate="82"
                     /db_xref="taxon:31647"
     gene            1. .540
                     /gene="NS3"
     CDS             <1. .>540
                     /gene="NS3"
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="BAB88284.1"
                     /db_xref="GI:19911949"
                     /translation="APITAYAQQTRGLLGCIVTSLTGRDKNQVEGEVQVVSTATQSFL
                     ATCINGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWPAPSGARSLTPCTCGSS
                     DLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPAGHAVGIFRAAVCT
                     RGVAKAVDFVPVESMATTMR"
     mat_peptide     1. .>540
                     /gene="NS3"
                     /product="nonstructural protein"
BASE COUNT       96 a    175 c    167 g    102 t
ORIGIN

Alignment Scores:
Pred. No.:                4.14          Length:          540
Score:                    43.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              97.73%        Indels:          0
DB:                       14            Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AB072102 (1-540)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||||.| |||||| |:::|:|||| :
Db         37 CTACTCGGCTGCATCGTCACTAGCCTC 63

RESULT 5
AB089533
LOCUS       AB089533                 540 bp    RNA     linear   VRL 02-APR-2003
DEFINITION  Hepatitis C virus NS3 gene for polyprotein, partial cds,
            isolate:Y-031.
ACCESSION   AB089533
VERSION     AB089533.1  GI:29467273
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1
  AUTHORS   Ogata,S. and Hotta,H.
  TITLE     Identification of Hepatitis C Virus (HCV) Subtype 1b Strains That
            Are Highly, or Only Weakly, Associated with Hepatocellular
            Carcinoma on the Basis of Secondary Structure of an Amino-Terminal
            Portion of the HCV NS3 Protein
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 540)
  AUTHORS   Hotta,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-AUG-2002) Hak Hotta, Kobe University Graduate School
            of Medicine, Division of Microbiology; Chuo-ku, Kusunoki-cho 7-5-1,
            Kobe, Hyogo 650-0017, Japan (E-mail:hotta@kobe-u.ac.jp,
            Tel:81-78-382-5500, Fax:81-78-382-5519)
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /isolate="Y-031"
                     /db_xref="taxon:11103"
                     /note="subtype: 1b"
     gene            1. .540
                     /gene="NS3"
     CDS             <1. .>540
                     /gene="NS3"
```

```
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="BAC67082.1"
                     /db_xref="GI:29467274"
                     /translation="APITAYAQQTRGLLGCIVTSLTGRDKNQVEGEVQVVSTATQSFL
                     ATCINGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSS
                     DLYLVTRHADVIPVRRRGDGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCT
                     RGVAKAVDFIPVESMETTMR"
BASE COUNT       96 a    176 c    166 g    102 t
ORIGIN

Alignment Scores:
Pred. No.:                4.14          Length:          540
Score:                    43.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              97.73%        Indels:          0
DB:                       14            Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AB089539 (1-540)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||:||||:|||||:::|||||||||
Db         37 CTACTTGGGTGTATCGTCACTAGCCTC 63

RESULT 6
AB089539
LOCUS       AB089539                 540 bp    RNA     linear   VRL 02-APR-2003
DEFINITION  Hepatitis C virus NS3 gene for polyprotein, partial cds,
            isolate:Y-038.
ACCESSION   AB089539
VERSION     AB089539.1  GI:29467285
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1
  AUTHORS   Ogata,S. and Hotta,H.
  TITLE     Identification of Hepatitis C Virus (HCV) Subtype 1b Strains That
            Are Highly, or Only Weakly, Associated with Hepatocellular
            Carcinoma on the Basis of Secondary Structure of an Amino-Terminal
            Portion of the HCV NS3 Protein
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 540)
  AUTHORS   Hotta,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-AUG-2002) Hak Hotta, Kobe University Graduate School
            of Medicine, Division of Microbiology; Chuo-ku, Kusunoki-cho 7-5-1,
            Kobe, Hyogo 650-0017, Japan (E-mail:hotta@kobe-u.ac.jp,
            Tel:81-78-382-5500, Fax:81-78-382-5519)
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /isolate="Y-038"
                     /db_xref="taxon:11103"
                     /note="subtype: 1b"
     gene            1. .540
                     /gene="NS3"
     CDS             <1. .>540
                     /gene="NS3"
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="BAC67083.1"
                     /db_xref="GI:29467286"
                     /translation="APITAYSQQTRGLLGCVITSLTGRDKNQVEGEVQVVSTATQSFL
                     ATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSS
                     DLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPLGHVVGIFRAAVCT
                     RGVAKAVDFVPVESMETTMR"
BASE COUNT       93 a    159 c    167 g    121 t
ORIGIN

Alignment Scores:
Pred. No.:                4.14          Length:          540
Score:                    43.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              97.73%        Indels:          0
DB:                       14            Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AB089539 (1-540)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||:||||:|||||:::|||||||||
Db         37 CTACTTGGGTGTATCGTCACTAGCCTC 63

RESULT 7
AB089558
LOCUS       AB089558                 540 bp    RNA     linear   VRL 02-APR-2003
DEFINITION  Hepatitis C virus NS3 gene for polyprotein, partial cds,
            isolate:Y-063.
ACCESSION   AB089558
VERSION     AB089558.1  GI:29467323
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1
  AUTHORS   Ogata,S. and Hotta,H.
  TITLE     Identification of Hepatitis C Virus (HCV) Subtype 1b Strains That
            Are Highly, or Only Weakly, Associated with Hepatocellular
            Carcinoma on the Basis of Secondary Structure of an Amino-Terminal
            Portion of the HCV NS3 Protein
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 540)
  AUTHORS   Hotta,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-AUG-2002) Hak Hotta, Kobe University Graduate School
            of Medicine, Division of Microbiology; Chuo-ku, Kusunoki-cho 7-5-1,
            Kobe, Hyogo 650-0017, Japan (E-mail:hotta@kobe-u.ac.jp,
            Tel:81-78-382-5500, Fax:81-78-382-5519)
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /isolate="Y-063"
                     /db_xref="taxon:11103"
                     /note="subtype: 1b"
     gene            1. .540
                     /gene="NS3"
     CDS             <1. .>540
                     /gene="NS3"
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="BAC67107.1"
                     /db_xref="GI:29467324"
                     /translation="APITAYSQQTRGLLGCIVTSLTGRDKNQVEGEVQVVSTATQSFL
                     ATCINGVCWTVYHGAGSKTLAGPKGPITQMYTNVDLDLVGWQAPPGARSMTPCTCGSS
                     DLYLVTRHADVIPVRRRSDGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCT
                     RGVAKAVDFVPVESMETTMR"
BASE COUNT       95 a    168 c    163 g    114 t
ORIGIN

Alignment Scores:
Pred. No.:                4.14          Length:          540
Score:                    43.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              97.73%        Indels:          0
DB:                       14            Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AB089558 (1-540)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
```

```
                    ||||||:|:|||||::|||| | |||
Db         37 CTACTTGGTTGCATCGTCACTAGCCTC 63

RESULT 8
AB100809
LOCUS       AB100809                 540 bp    RNA     linear   VRL 02-APR-2003
DEFINITION  Hepatitis C virus NS3 gene for polyprotein, partial cds, isolate:
            Y-CH-06-B.
ACCESSION   AB100809
VERSION     AB100809.1  GI:29467656
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1
  AUTHORS   Ogata,S. and Hotta,H.
  TITLE     Identification of Hepatitis C Virus (HCV) Subtype 1b Strains That
            Are Highly, or Only Weakly, Associated with Hepatocellular
            Carcinoma on the Basis of Secondary Structure of an Amino-Terminal
            Portion of the HCV NS3 Protein
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 540)
  AUTHORS   Hotta,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2003) Hak Hotta, Kobe University Graduate School
            of Medicine, Division of Microbiology; Chuo-ku, Kusunoki-cho 7-5-1,
            Kobe, Hyogo 650-0017, Japan (E-mail:hotta@kobe-u.ac.jp,
            Tel:81-78-382-5500, Fax:81-78-382-5519)
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /isolate="Y-CH-06-B"
                     /db_xref="taxon:11103"
                     /note="subtype: 1b"
     gene            1. .540
                     /gene="NS3"
     CDS             <1. .>540
                     /gene="NS3"
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="BAC67219.1"
                     /db_xref="GI:29467657"
                     /translation="APITAYSQQTRGLLGCVITSLTGRDKNQVEGEVQVVSTAIQSFL
                     AICVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSS
                     DLYLVTRHADVIPVRRRGDGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCT
                     RGVAKAVDFVPVESMETTMR"
BASE COUNT       93 a    171 c    166 g    110 t
ORIGIN

Alignment Scores:
Pred. No.:                4.14          Length:          540
Score:                    43.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              97.73%        Indels:          0
DB:                       14            Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AB100809 (1-540)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||||||||||||||:|||||||||||
Db         37 CTACTTGGTTGCATCGTCACTAGCCTC 63

RESULT 9
AB100827
LOCUS       AB100827                 540 bp    RNA     linear   VRL 02-APR-2003
DEFINITION  Hepatitis C virus NS3 gene for polyprotein, partial cds, isolate:
            Y-HCC-15-B.
ACCESSION   AB100827
VERSION     AB100827.1  GI:29467692
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1
  AUTHORS   Ogata,S. and Hotta,H.
  TITLE     Identification of Hepatitis C Virus (HCV) Subtype 1b Strains That
            Are Highly, or Only Weakly, Associated with Hepatocellular
            Carcinoma on the Basis of Secondary Structure of an Amino-Terminal
            Portion of the HCV NS3 Protein
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 540)
  AUTHORS   Hotta,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2003) Hak Hotta, Kobe University Graduate School
            of Medicine, Division of Microbiology; Chuo-ku, Kusunoki-cho 7-5-1,
            Kobe, Hyogo 650-0017, Japan (E-mail:hotta@kobe-u.ac.jp,
            Tel:81-78-382-5500, Fax:81-78-382-5519)
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /isolate="Y-HCC-15-B"
                     /db_xref="taxon:11103"
                     /note="subtype: 1b"
     gene            1. .540
                     /gene="NS3"
     CDS             <1. .>540
                     /gene="NS3"
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="BAC67237.1"
                     /db_xref="GI:29467693"
                     /translation="APITAYSQQTRGLLGCVITSLTGRDKNQVEGEVQVVSIATQSFL
                     AICVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSS
                     DLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPLGHVVGIFRAAVCT
                     RGVAKAVDFVPVESMETTMR"
BASE COUNT       93 a    162 c    166 g    119 t
ORIGIN

Alignment Scores:
Pred. No.:                4.14          Length:          540
Score:                    43.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              97.73%        Indels:          0
DB:                       14            Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AB100827 (1-540)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||||:|||||||:||:|||||||||
Db         37 CTACTTGGCTGCGTTATCACTAGCCTC 63

RESULT 10
AY070171
LOCUS       AY070171                 1211 bp    RNA     linear   VRL 01-APR-2002
DEFINITION  Hepatitis C virus strain 687 polyprotein gene, partial cds.
ACCESSION   AY070171
VERSION     AY070171.1  GI:19879463
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 1211)
  AUTHORS   Kalinina,O., Norder,H., Mukomolov,S. and Magnius,L.O.
  TITLE     A natural intergenotypic recombinant of hepatitis C virus
            identified in St. Petersburg
  JOURNAL   J. Virol. 76 (8), 4034-4043 (2002)
  MEDLINE   21904745
    PUBMED  11907242
REFERENCE   2  (bases 1 to 1211)
```

```
  AUTHORS   Kalinina,O., Norder,H., Mukomolov,S. and Magnius,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2001) Virology, Swedish Institue for Infectious
            Disease Control, Solna SE 171 82, Sweden
FEATURES             Location/Qualifiers
     source          1. .1211
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /strain="687"
                     /db_xref="taxon:11103"
                     /note="recombinant 2k/1b"
     CDS             <1. .>1211
                     /note="contains E2, p7, NS2 and NS3"
                     /codon_start=3
                     /product="polyprotein"
                     /protein_id="AAL58587.1"
                     /db_xref="GI:19879464"
                     /translation="EWVVLLFLLLADARVCACLWMLLLGQAEAALEKLVILHAASAA
                     SFNGIICFTIFFIAAWHIKGRVVPLATYSYIGLWSFSLLLALPQQAYALNTTEGCGL
                     GLVLLVVISVLTLSPAYKTLLGRSLWWLSYLLVRAEALIQEWVPPLQARGGRDGIIWA
                     AVILCPGIVFDITKWLLAVLGPGHLIRSVLTSVPYFVRAQALLRICAAVRHLSGGKYV
                     QMMLLTLGRWILTYIYDHLTPLQDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAA
                     CGDIILGLPVSARRGREILLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIVTSLTGR
                     DRNQVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDL
                     VGWPAPPGARS"
BASE COUNT      225 a    358 c    349 g    279 t
ORIGIN

Alignment Scores:
Pred. No.:              9.45          Length:         1211
Score:                  43.00         Matches:        8
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  88.89%        Mismatches:     0
Query Match:            97.73%        Indels:         0
DB:                     14            Gaps:           0

US-09-965-594-1_COPY_14_22 (1-9) x AY070171 (1-1211)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||||||||||||||||:|||||| |||
Db        969 CTACTTGGCTGCATCGTCACTAGCCTC 995

RESULT 11
AF187824/c
LOCUS       AF187824               2691 bp    RNA     linear   VRL 02-MAY-2001
DEFINITION  Cache Valley virus RU68 G1 gene, partial cds.
ACCESSION   AF187824
VERSION     AF187824.1  GI:6959495
KEYWORDS    .
SOURCE      Cache Valley virus
  ORGANISM  Cache Valley virus
            Viruses; ssRNA negative-strand viruses; Bunyaviridae;
            Orthobunyavirus; Bunyamwera virus group.
REFERENCE   1  (bases 1 to 2691)
  AUTHORS   Brockus,C.L. and Grimstad,P.R.
  TITLE     Comparative analysis of G1 glycoprotein-coding sequences of Cache
            Valley virus (Bunyaviridae: Bunyavirus) isolates
  JOURNAL   Virus Genes 22 (2), 133-139 (2001)
  MEDLINE   21221452
   PUBMED   11324749
REFERENCE   2  (bases 1 to 2691)
  AUTHORS   Brockus,C.L., Collins,F.H., Besansky,N.J. and Grimstad,P.R.
  TITLE     Phylogenetic Analysis of Cache Valley Virus Isolates
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 2691)
  AUTHORS   Brockus,C.L. and Grimstad,P.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-1999) Vector Biology Labs, University of Notre
            Dame, Galvin Life Sciences Center, Notre Dame, IN 46556, USA
FEATURES             Location/Qualifiers
     source          1. .2691
                     /organism="Cache Valley virus"
                     /mol_type="genomic RNA"
                     /strain="RU68"
                     /db_xref="taxon:80935"
                     /country="USA: Dennisville, NJ"
                     /note="1982 isolate from Ae. sollicitans"
     CDS             <1. .>2691
                     /note="envelope glycoprotein; putative soluble portion"
                     /codon_start=1
                     /product="G1"
                     /protein_id="AAF33116.1"
                     /db_xref="GI:6959496"
                     /translation="ETDCWTKKSLEIDCIGPMLDIGSCTNKNTRSYDAEAQKLVSQSK
                     ISQLDADQVSLLGNSIDSAIKAIRAQKTYPTMHLLETIFLMKNCDYYTSFNHNSGYSQ
                     AKWRLIAKTGHFDICSRHSTHHFCRCINDGTKCQNGNWDFAGEMNETYNSKNDFYSHD
                     GNLFYIIFENAFPGITESLFYQLIQKKNTTGASKVLGKLITKYGNNNMFVGVWKFGQY
                     LLTLPYINNTELSQNRRLTILATQELLSRSHSGKQESMSNAIQGSITKECHNAKGVGC
                     ISPRFGIPIGNLTACGDSPNYKIYKTPEKLYKSNNKGFVWCSNDVHCLNDFEPADEEV
                     VNKTKKLTCFLTDVSVEVDIFSVAASTCKMADKGVCTVNGAKWNIIKCDSGLYYFTDH
                     REGQDTGNDFGEYCTSHSCNTERFPINPDILKDCTWEFHSRKSKYISTISLESLEEFR
                     RAMAEKLSHTLTIYNFKPTANLPHIRPTYKFITANGVENSDGIESAYLLSSIPALGGI
                     SDGYDVQSKDSFPLLDMIVFIDSAVIRTTYYHIYDTGHTIGIYTQHDEHCTGPYPAMV
                     PHKENWMTFSQERTSRWGCEEFGCLAINTGCVFGSCQDIIHPETKVYRKAVEEEVVLT
                     VCINFPGNNYCTEINAIEPKITDEIELQFKTVDTKTLPNILAVQNHKLYSGQINDLGS
                     FSQGCGNVQKTNHSILGMGTVKFDYVCHCASRKDVIVRRCYNNNFDSCRLLKEETSLI
                     FADNHETLEVAHTKHLIGELQFKIMLGDIRYKSFAESPELEIDAKCVGCPSCFESYSC
                     NFQIVTNIDTVCSVEGPCTLFHNRIISANKQSYGLKMSCQTKPNQNEEFIICNRKYS
                     VLFTTTDKNDKIEVNTGDQTSYIYEKDSRCKTWLCRVRDEG"
BASE COUNT      949 a    476 c    539 g    727 t
ORIGIN

Alignment Scores:
Pred. No.:              21.3          Length:         2691
Score:                  43.00         Matches:        8
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  88.89%        Mismatches:     0
Query Match:            97.73%        Indels:         0
DB:                     14            Gaps:           0

US-09-965-594-1_COPY_14_22 (1-9) x AF187824 (1-2691)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||||||:||||:::|||:|||||||
Db       1387 TTGTTAGGGTGTGTGATAACTTCTCTG 1361

RESULT 12
AF231116/c
LOCUS       AF231116               2691 bp    DNA     linear   VRL 02-MAY-2001
DEFINITION  Cache Valley virus isolate Ar78-5912 G1 envelope glycoprotein gene,
            partial cds.
ACCESSION   AF231116
VERSION     AF231116.1  GI:13249094
KEYWORDS    .
SOURCE      Cache Valley virus
  ORGANISM  Cache Valley virus
            Viruses; ssRNA negative-strand viruses; Bunyaviridae;
            Orthobunyavirus; Bunyamwera virus group.
REFERENCE   1  (bases 1 to 2691)
  AUTHORS   Brockus,C.L. and Grimstad,P.R.
  TITLE     Comparative analysis of G1 glycoprotein-coding sequences of Cache
            Valley virus (Bunyaviridae: Bunyavirus) isolates
  JOURNAL   Virus Genes 22 (2), 133-139 (2001)
  MEDLINE   21221452
   PUBMED   11324749
REFERENCE   2  (bases 1 to 2691)
  AUTHORS   Brockus,C.L. and Grimstad,P.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2000) Biological Sciences, University of Notre
            Dame, Galvin Life Sciences Bldg, Notre Dame, IN 46556, USA
FEATURES             Location/Qualifiers
     source          1. .2691
                     /organism="Cache Valley virus"
                     /mol_type="genomic DNA"
                     /isolate="Ar78-5912"
                     /specific_host="Anopheles quadrimaculatus"
```

```
                     /db_xref="taxon:80935"
                     /country="USA: Ohio, Kent"
     CDS             <1..>2691
                     /codon_start=1
                     /product="G1 envelope glycoprotein"
                     /protein_id="AAK16673.1"
                     /db_xref="GI:13249095"
                     /translation="ETDCWTKKSLDIDCIGPMLDIGSCTNKNTRSYDAEAQKLVSQSK
                     ISQLDADQVSLLGNSIDSAIKAIRAQKTYPTMHLLETIFLMKNCDYYTSFNHNSGYSQ
                     AKWRLIAKTGHFDICSRHSTHHFCRCINDGTKCQNGNWDFAGEMNETYNSKNDFYSHD
                     LNLFYTIFENAFPGTTESLFYQLIQKKNTTGASKVLGKLITKYGNNNMFVGVWKFGQY
                     LLTLPYTNNTELSQNRRLTILATQELLSRSHSGRQESMSNAIQGSITKECHNAKGVGC
                     ISPRFGIPIGNLTACGDSPNYKIYKTPEKLYKSNNKGEVWCSNDVHCLNDFEPADEEV
                     VNKIKKLTCFLTDVSVEVDIFSVAASTCKMADKGVCTVNGAKWNIIKCDSGLYYFTDH
                     REGQDTGNDFGEYCISHSCNTERFPINPDILKDCTWEFHSRKSKYISTISLESLEEFR
                     RAMAEKLSHTLTIYNFKPTANLPHIRPTYKFITANGVENSDCIESAYLLSSIPALGGI
                     SVGYNVQSKDNFPLLDMIVFINSAVIRTTYYHIYDIGHTIGIYTQHDEHCTGPCPAVV
                     PHKENWMTFSQERTSRWGCEEFGCLAINTGCVFGSCQDIIHPETKVYRKAVEEEVVLT
                     VCINFPGNNYCTEINAIEPKITDEIELQFKTVDTKILPNLLAVQNHKLYSGQINDLGS
                     FSQGCGNVQKTNHSILGMGTVKFDYVCHGASRKDVIVRRCYNNNFDSCKLLKEETSLI
                     FADNHETLEVAHTKHLIGELQFKIMLGDIRYKSFAESPELEIDAKCVGCPSCFESYSC
                     NFQIVTNIDTVCSVEGPCTLFHNRIIISANKQSYGLKMSCQTKPNQNEEFIICNRKYS
                     VLFTTIDKNDKIEVNTGDQTSYIYEKDSRCKTWLCRVRDEG"
BASE COUNT      947 a    475 c    541 g    728 t
ORIGIN

Alignment Scores:
Pred. No.:                21.3          Length:          2691
Score:                    43.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              97.73%        Indels:          0
DB:                       14            Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AF231118 (1-2691)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||:|.|||||:::|||| |||| ||
Db       1387 TTGTTAGGGTGTGTGATAACTTCTCTG 1361

RESULT 13
AF231118/c
LOCUS       AF231118                2691 bp    DNA     linear   VRL 02-MAY-2001
DEFINITION  Cache Valley virus isolate 61D240 G1 envelope glycoprotein gene,
            partial cds.
ACCESSION   AF231118
VERSION     AF231118.1  GI:13249098
KEYWORDS    .
SOURCE      Cache Valley virus
  ORGANISM  Cache Valley virus
            Viruses; ssRNA negative-strand viruses; Bunyaviridae;
            Orthobunyavirus; Bunyamwera virus group.
REFERENCE   1  (bases 1 to 2691)
  AUTHORS   Brockus,C.L. and Grimstad,P.R.
  TITLE     Comparative analysis of G1 glycoprotein-coding sequences of Cache
            Valley virus (Bunyaviridae: Bunyavirus) isolates
  JOURNAL   Virus Genes 22 (2), 133-139 (2001)
  MEDLINE   21221452
   PUBMED   11324749
REFERENCE   2  (bases 1 to 2691)
  AUTHORS   Brockus,C.L. and Grimstad,P.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2000) Biological Sciences, University of Notre
            Dame, Galvin Life Sciences Bldg, Notre Dame, IN 46556, USA
FEATURES             Location/Qualifiers
     source          1..2691
                     /organism="Cache Valley virus"
                     /mol_type="genomic DNA"
                     /isolate="61D240"
                     /specific_host="Psorophora cingulata"
                     /db_xref="taxon:80935"
                     /country="Mexico: Bocas el Toro"
                     /note="isolated in 1961
                     subtype: Tlacotalpan"
     CDS             <1..>2691
                     /codon_start=1
                     /product="G1 envelope glycoprotein"
                     /protein_id="AAK16675.1"
                     /db_xref="GI:13249099"
                     /translation="GTDCWTKKSLDIDCIGPMLDIGSCTNKNTRNYDIEAQKLVSQSK
                     VSQLDADQASLLGNSIDSAIKAIRAQKTYPTMHLLETIFLMKNCDYYTSFNHNSGYSQ
                     AKWRLIAKTGHFDICSRHSTHHFCRCINDGTKCQNGDWDFAGEMNETYNSKNDFYSHD
                     LNLFYTIFENAFPGTTESLFYQLIQKKNTTGASKILGKLITKYGNKNMFVGVWKFGQY
                     LLTLPYINNTELSQNRRQTILATQESLSRSHSGRQESMSSAIQGSVTKECHNAKDVGC
                     ISPRFGIPIGNLTACGDSPNYKIYKTPEKLYKSNNKGEVWCSNDVHCLNDFEPADEEV
                     VNKVKKLTCFLTDVSVKTDIFSVAASICKMADKGVCTVNDAKWNIIKCDSGLYYFTDH
                     REGQDTGNDFGEYCISHSCNTERFPINPDILKDCTWEFHSRKSKYISTISLESLEEFR
                     RAMTEKLSHTLTIYNFKPTANLPHIRPTYKFITANGVENSDGIESAYILSSIPALGGI
                     SVGYNVQTKDNFPLLDIIVFIKSAVIRTTYNHIYDTGPTIGINTQHDEHCTGPCPAVV
                     PHKENWITFSQERTSRWGCEEFGCLAINTGCVFGSCQDIIHPETKVYRKAVEEEVILI
                     VCINFPGNNYCTEINAIEPKITDEIELQFKTVDTKILPNLLAVQNHKLYSGQINDLGS
                     FSQGCGNVQKINHSILGMGTAKFDYVCHGASRKDVIVRRCYNNNFDSCKLLKEETSLI
                     FADNHETLEVAHTKHLIGELQFKIMLGDIRYKSFAESPELEIDAKCVGCPSCFESYSC
                     NFQIVTNIDTVCSVEGPCTLFHNRIIISANKQSYGLKMSCQTKPNQNEEFIICNRKYS
                     VLFTTIDKNDKIEVNTGDQTSYIYEKDSRCKTWLCRVRDEG"
BASE COUNT      962 a    468 c    535 g    726 t
ORIGIN

Alignment Scores:
Pred. No.:                21.3          Length:          2691
Score:                    43.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              97.73%        Indels:          0
DB:                       14            Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AF231118 (1-2691)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||||||||||||:::||'| |||| ||
Db       1387 TTGTTAGGGTGTGTGATAACTTCTCTG 1361

RESULT 14
AF186241/c
LOCUS       AF186241                4463 bp    RNA     linear   VRL 02-MAY-2001
DEFINITION  Cache Valley virus strain MI80-1-450 M segment, complete sequence.
ACCESSION   AF186241
VERSION     AF186241.1  GI:6959489
KEYWORDS    .
SOURCE      Cache Valley virus
  ORGANISM  Cache Valley virus
            Viruses; ssRNA negative-strand viruses; Bunyaviridae;
            Orthobunyavirus; Bunyamwera virus group.
REFERENCE   1  (bases 1 to 4463)
  AUTHORS   Brockus,C.L. and Grimstad,P.R.
  TITLE     Comparative analysis of G1 glycoprotein-coding sequences of Cache
            Valley virus (Bunyaviridae: Bunyavirus) isolates
  JOURNAL   Virus Genes 22 (2), 133-139 (2001)
  MEDLINE   21221452
   PUBMED   11324749
REFERENCE   2  (bases 1 to 4463)
  AUTHORS   Brockus,C.L., Collins,F.H., Besansky,N.L. and Grimstad,P.R.
  TITLE     Phylogenetic analysis of Cache Valley virus isolates
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 4463)
  AUTHORS   Brockus,C.L. and Grimstad,P.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-SEP-1999) Vector Biology Labs, University of Notre
            Dame, Galvin Life Sciences Center, Notre Dame, IN 46556, USA
FEATURES             Location/Qualifiers
     source          1..4463
                     /organism="Cache Valley virus"
                     /mol_type="genomic RNA"
                     /strain="MI80-1-450"
                     /db_xref="taxon:80935"
                     /chromosome="M segment"
```

```
                     /country="USA: Cass county, Michigan"
                     /note="1981 isolate from equid"
     misc_feature    1. .48
                     /note="signal sequence"
     CDS             49. .4353
                     /note="G2-M-G1"
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAF33113.1"
                     /db_xref="GI:6959490"
                     /translation="MLPLLILAFLGFSTSNFIATRQFHQGHLIAFSKSAISISSPCLK
                     DDVSMLKSEVVYTKNDTGLFGHNKVFRSWTIKDWKLCNPIPIAGGTINVIEVNKDLSL
                     MTKTYICSRDCTIIDKEEAQIIFVTDKLNHFEVSGTILSSGWFKSKASVTLDRTCEH
                     VKVICGKKSLQFHACFKHHMSCVRFFHNTILPGNMITSMCQNIELIIILGLTLAIFIL
                     MVILTKTYICYLLMPIFMPIAYLYGWAYNKSCKKCNCCGLAYHPFTNCGSHCVCGLKF
                     EASDRMRIHRESGLCQGYKSLRIARLLCKSKGSSLVISGLLSMLILSFVTPIHGTITS
                     YPESRKYTLEEITDVVEGLVSEQNIKEYIVFYTSLFGSLILLMAFIMSITLSKITSIL
                     AKRNVIYCEECSMYHSKRGIKYNGDFTNKGFCTCNEQEDTEGLVTHKVSKRCTYKYQ
                     LKWAKVLMITLVLLLIAQNTILIVAAETDCWTKKSLDIDCIGPMLDIGSCTNKNTRSY
                     DAEAQKLVSQSKISQLDADQVSLDGNSIDSAIKAIRAQKTYFTMHLLETIPIMKNCDY
                     YTSFNHNSGYSCAKWRLIAKTDHFDTCSRHSTHHPCRCINDGTKCQNSNWDFAGEMNE
                     TYNSKNDFYSHDLNLPYTIPENAFPGTIESLFYDLDCKKNTTGASQVLGKLITKYCNK
                     NMFVGVWKPGQYLLTLPYINNTELSQNRRLTLAIQELLSRSHSGRQFSMSNAIQCST
                     TKECHNAKGVGCISPRFGIPIGNITAGSDSPNYKIYKTPEKLYKSNNKGEVWCSNDVH
                     CLNDFEPADEEVVNKIKKLTCFLTDVSVFVDIFSVAASTCKADKSVCTVNSAKWNLI
                     KCDSGLYYPTLHRESQDTGNDPGEYCISHSCNTDRFPINFDILKDCTWEPHSRKSKYI
                     STISLESLEEFRRAMAEKLSHTLTIYNFKFTANLPHIRPIYKFITANGVENSDGIRSA
                     YILSSIPALGGTSVGYNVQTKDNFPLLDIIVFIKSAVIRTTYNHIYETGPTIGINTQH
                     DEHCTGPCPAVVPHKENWMTFSQERTSEWGCEEPGCLAINTGCVPQSCQDITHPLTKV
                     YRKAVEEEVVLTVCINFPGNNYCTEINAIFPKIILEIELQFKTVDTKTLPNLLAVQNH
                     KLYSGQINDLGSFSQGCGNVQKTNHSILGMGTVKFDYVCHGASKKDVIVRRCYNNNFD
                     SCKLLKEETSLIFADNHETLEVAHTKHLIGFLQFKIMLGDIRYKSFAESPKLEIDAKC
                     VGCPSCFESYSCNFQIVTNIDIVCSVEGPCTLFHNRIILSANKQSYGLKMSCQTKPNQ
                     NEEFIICNRKYSVLFTPIDKNDKIEVNTGDQTSYIYKKDSRCKTWLCRVFPEGISVLF
                     EPLKAFFGSYFSIIFYVVVGIIVLFLIIYIFLPMFFKLKDVLKRNEYLYLQFLKHK"
     mat_peptide     49. .951
                     /product="G2"
                     /note="envelope glycoprotein"
     mat_peptide     952. .1476
                     /product="NSm"
                     /note="non-structural protein"
     mat_peptide     1477. .4350
                     /product="G1"
                     /note="envelope glycoprotein"
BASE COUNT     1550 a     756 c     869 g    1288 t
ORIGIN

Alignment Scores:
Pred. No.:                35.7          Length:          4463
Score:                    43.00         Matches:         9
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              97.73%        Indels:          0
DB:                       14            Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AF186241 (1-4463)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||:|||||:|||:::|||:|:||||||
Db       2863 TTGTTAGGGTGTGTGATAACTTCTCTG 2837

RESULT 15
AF313916
LOCUS       AF313916                9359 bp    mRNA    linear   VRL 01-JAN-2002
DEFINITION  Hepatitis C virus polyprotein mRNA, partial cds.
ACCESSION   AF313916
VERSION     AF313916.1  GI:18027684
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 9359)
  AUTHORS   Fanning,L.J., Itakura,J., Nagayama,K. and Enomoto,N.
  TITLE     Characteristics of Hepatitis C viral genome associated with disease
            progression in a homogeneous patient population
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9359)
  AUTHORS   Fanning,L.J., Itakura,J., Nagayama,K. and Enomoto,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-2000) Medicine, National University of Ireland,
            Cork, Hepatitis C Unit, Clinical Sciences Building, Cork University
            Hospital, Cork, Ireland
FEATURES             Location/Qualifiers
     source          1. .9359
                     /organism="Hepatitis C virus"
                     /mol_type="mRNA"
                     /db_xref="taxon:11103"
                     /note="isolated from source plasma identified as the
                     single source of the iatrogenic infection of Rhesus
                     negative women in Ireland from May 1977-November 1978;
                     exposure was through HCV 1b contaminated anti-D
                     immunoglobulin"
     5'UTR           1. .329
     CDS             330. .>9359
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAL55821.1"
                     /db_xref="GI:18027685"
                     /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKARRPEGRTWAQPGYPWPLYGNEGMGWAGWLLSPRG
                     SRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLED
                     GVNYATGNLPGCSFSIFLLALLSCLTIPASAIEVRNVSGVYHVTNDCSNASIVYEAAD
                     MIMHTPGCVPCVRENNSSRCWVALTPTLAARNSSIPTTTIRRHVDLLVGTAAFCSAMY
                     VGDLCGSVFLVSQLFTFSPRRHETVQDCNCSIYPGHVSGHRMAWDMMMNWSPTTALVV
                     SQLLRIPQAVVDMVAGAHWGVLAGLAYYSMAGNWAKVLIVMLLFAGVDGETYTTGGAQ
                     AHTTRGLAILFSFGPSQNIQLINTNGSWHINRTALNCNDSLNTGFLAALFYAHRFNAS
                     GCPERMASCRPIDKFAQGWGPIAYAKPLSLDQKPYCWHYAPQPCGIVPASQVCGPVYC
                     FTPSPVVVGTTDRFGVPTYSWGENETDVLLLNNTRPPRGNWFGCTWMNGTGFTKTCGG
                     PPCDIGGVGNDTLICPTDCFRKHPEATYTKCGSGPWLTPRCMVDYPYRLWHYPCTVNF
                     TIFKVRMYVGGVEHRLNAACNWTRGERCDLEDRDRSELSPLLLSTTEWQVLPCSFTTL
                     PALSTGLIHLHRNIVDVQYLYGIGSAVVSYAIKWEYVLLLFLLADARVCACLWMMLL
                     IVQAEAALENLVVLNAASVAGTHGILSFLVFFCAAWYIKGRLVPGAAYAFYGVWPLLL
                     LLLALPPRAYAMDREMAASCGGAVFIGLALITLSPHYKVFLARLIWWLQYFITRAEAH
                     LQVWIPPLNVRGGRDAIILLMCAIHPELIFAITKILLAILGPLMVLQAGITRVPYFVR
                     AHGLIRACVLVREVAGGHYVQMALMRLAALTGTYVYDHLTPLRDWAHEGLRDLAVAVE
                     PVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADSLEGQGWRLLAPIT
                     AYAQQTRGLLGCIVTSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGTK
                     TLAGQKGPITQMYTNVDQDLVGWLAPPGARSLTPCTCGSSDLYLVTRHADVIPVRRRG
                     DSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMETT
                     MRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNPSVAATL
                     GFGAYMSKAHGVDPNIRIGVRTITTGAPITYSTYGKFLADGGCSGGAYDIIICDECHS
                     TDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNIGEIPFYGK
                     AIPIETIKGGRHLIFCHSRKKCDELAAKLSGLGLNAVAYYRGLDVSVIPTSGDVVVVA
                     TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRG
                     RTGIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQ
                     DHLEFWEGVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKC
                     LIRLKPTLHGPTPLLYRLGAVQNEVILTHPITKFIMACMSADLEVATSTWVLVGGVLA
                     ALAAYCLTTGSVVIVGRIILSGKPAIVPDREVLYQQFDEMEECATHLPYIEQGMQLAE
                     QFKQKALGLLQTATKQAEAAAPVVESKWRALESFWAKHMWNFISGIQYLAGLSTLPGN
                     PAIASLMAFTASITSPLTTQHTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSI
                     GLGKVLVDILAGYGAGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVVCAA
                     ILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLK
                     RLHQWINEDCSTPCSGSWLRDVWDWICTVLTDFKTWLQSKLLPRLPGVPFLSCQRGYK
                     GVWRGDGIMQTTCPCGAQITGHVKNGSMRIVGPRTCSNIWHGTFPINAYTTGPCTPSP
                     APNYSRALWRVTAEEYLEVTQVGDFHYVTGMTTDNVKCPCQVPAPEFFTEVDGVRLHR
                     YAPPCKPLLREEVTFQVGLNQYLVGSQLPCEPEPDVAVLTSMLTDPSHITAEAAKRRL
                     ARGSPPSLASSSASQLSAPSLKATCTTRHDSPDADLIEANLLWRQEMGGNITRVESEN
                     KVVILDSFEPLRAEEDEREVSVPAEILRRSRKFPRAMPIWARPDYNPPLVESWKDPDY
                     VPPVVHGCPLPPIKAPPIPPPRRKRTVVLTESTVSSALAELATKTFGSSESSAVDSGT
                     ATAPPDQPSDDGDAGSDVESYSSMPPLEGEPGDPDLSDGSWSTVSEEASEDVVCCSMS
                     YTWTGALITPCSAEESKLPINALSNSLLRHHNMVYATTSRSASQRQKKVTFDRLQVLD
                     DHYRDVLKEMKAKASTVKAKLLPVEEACKLTPPHSAKSKFGYGAKDVRNLSSKAVNHI
                     LSVWKDLLEDTETPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYD
                     VVSTLPQAVMGSSYGFQYSPGQRVEFLVNAWKKKKNPMGFAYDTRCFDSTVIESDIRV
                     EESIYQCCDLAPEARQAIRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTL
```

TCYLKASAACRAAKLQDCIMLVCGDDLVVICESAGIXEDAASLRVFTEAMIRYSAPPG
DPPKPEYDLELITSCSSNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWL
GNIIMYAPTLWARMILMTHFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIQRLH
GLSAFSLRSYSPGEINRVAACLRKLGVPPLRVWRHRARSVRAKLLSQGGRAANCGKYL
FNWAVRTKLKLTPIPAASQLDLSSWFVAGYSGGDIYHSLSRARPRWFMLCLLLLSVGV
GIYLLPNR"

BASE COUNT 1877 a 2830 c 2680 g 1968 t 4 others
ORIGIN

Alignment Scores:

Pred. No.:	76	Length:	9359
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	97.73%	Indels:	0
DB:	14	Gaps:	0

US-09-965-594-1_COPY_14_22 (1-9) x AF313916 (1-9359)

```
Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||||:|||||||||:::|||||:|||
Db       3444 CTACTTGGCTGCATCG:CACCAGCCTC 3470
```

Search completed: September 30, 2003, 12:51:59
Job time : 2207 secs

GenCore version 5.1.6

CM protein - nucleic search, using frame_plus_p2n model

Run on: September 29, 2003, 19:09:03 ; Search time 189 Seconds
(without alignments)
128.545 Million cell updates/sec

Title: US-09-965-594-1_COPY_14_22
Perfect score: 44
Sequence: 1 LLGCIITSL 9

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105271

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 99%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x11
-Q=/cgn2_1/USPTO_spool/US09965594/runat_29092003_164249_14064/app_query.fasta_1.199
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=99.9 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594_@CGN_1_1_312_@runat_29092003_164249_14064 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		43	97.7	9436	15	AAQ63499	Blood transmiscibl
2		42	95.5	819	15	AAQ58817	NANBH virus gene f
3		42	95.5	3461	15	AAQ64068	Non-A, non-B hepat
4		42	95.5	3461	16	AAT30386	5'UTR/CORE/ENV/NS1
5	c	41	93.2	4977	23	AAS86668	DNA encoding novel
6	c	41	93.2	4977	23	AAS88891	DNA encoding novel
7	c	41	93.2	4977	23	AAS89842	DNA encoding novel
8		41	93.2	9595	20	AAX24843	Infectious hepatit
9		41	93.2	9595	22	AAC86939	Nucleotide sequenc
10		41	93.2	9595	22	AAF23492	Infectious Hepatit
11		41	93.2	94895	21	AAF22302	BAC containing rep
12	c	41	93.2	94895	21	AAF22302	BAC containing rep
13		41	93.2	103929	21	AAF22287	BAC containing rep
14		41	93.2	134499	21	AAF22286	BAC containing rep
15		41	93.2	1082138	21	AAF22305	Arabidopsis thalia
16	c	41	93.2	1082138	21	AAF22305	Arabidopsis thalia
17		40	90.9	424	24	ABL63442	Breast cancer rela
18		40	90.9	1933	20	AAX23258	HCV NS3 DNA. Hepa
19	c	40	90.9	2346	24	AAS19166	cDNA encoding huma
20	c	40	90.9	2520	25	ABX05429	Human novel polynu
21		40	90.9	8145	20	AAX23259	Plasmid pET-BS(+)/
22	c	40	90.9	910715	20	AAX20248	Borrelia burgdorfe
23	c	39	88.6	4154	25	ABT19487	Aspergillus fumiga
24		39	88.6	4746	21	AAZ45995	Rat calcium/calmod
25		39	88.6	9402	14	AAQ41345	Human hepatitis C
26		39	88.6	9405	14	AAQ40426	Full-length Hepati
27		38	86.4	242	25	ABX23888	Human GDP-mannose
28		37	84.1	39	20	AAX80326	HCV NS4A-NS3 compl
29	c	37	84.1	39	20	AAX80327	HCV NS4A-NS3 compl
30		37	84.1	39	20	AAX80328	HCV NS4A-NS3 compl
31	c	37	84.1	39	20	AAX80329	HCV NS4A-NS3 compl
32		37	84.1	39	20	AAX80332	HCV NS4A-NS3 compl
33	c	37	84.1	39	20	AAX80333	HCV NS4A-NS3 compl
34		37	84.1	39	20	AAX80334	HCV NS4A-NS3 compl
35	c	37	84.1	39	20	AAX80335	HCV NS4A-NS3 compl
36		37	84.1	39	20	AAX80316	HCV NS4A-NS3 compl
37	c	37	84.1	39	20	AAX80317	HCV NS4A-NS3 compl
38		37	84.1	39	20	AAX80318	HCV NS4A-NS3 compl
39	c	37	84.1	39	20	AAX80319	HCV NS4A-NS3 compl
40		37	84.1	39	20	AAX83816	HCV NS4A-NS3 compl
41	c	37	84.1	39	20	AAX83817	HCV NS4A-NS3 compl
42	c	37	84.1	39	20	AAX80305	HCV NS4A-NS3 compl
43		37	84.1	39	20	AAX80308	HCV NS4A-NS3 compl
44	c	37	84.1	39	20	AAX80309	HCV NS4A-NS3 compl
45		37	84.1	39	20	AAX80312	HCV NS4A-NS3 compl

ALIGNMENTS

RESULT 1
AAQ63499
ID AAQ63499 standard; cDNA; 9436 BP.
XX
AC AAQ63499;
XX
DT 17-JAN-1995 (first entry)
XX
DE Blood transmiscible NANBHV genome.
XX
KW Polymerase chain reaction; PCR; amplify; primer; non-A, non-B hepatitis;
KW NANBH; virus; blood transmissible; detection; hepatitis virus; RT-PCR;
KW C100 antibody; HCV RNA; NS5 region; ds.
XX
OS Non-A, non-B hepatitis virus.
XX
FH Key Location/Qualifiers

```
FT   CDS             342..9374
FT                   /*tag= a
XX
PN   JP06105690-A.
XX
PD   19-APR-1994.
XX
PF   10-MAR-1992;   92JP-0051885.
XX
PR   10-MAR-1992;   92JP-0051885.
XX
PA   (KAEN/) KAENNO K.
XX
DR   WPI; 1994-163130/20.
DR   P-PSDB; AAR53417.
XX
PT   Blood-transmissible non-A non-B hepatitis virus DNA - used for
PT   detection of hepatitis virus
XX
PS   Claim 1; Page 8-20; 22pp; Japanese.
XX
CC   This sequence represents the genome of a blood transmissible non-A,
CC   non-B hepatitis (NANBH) virus. This sequence was isolated using the
CC   primers given in AAQ63500-35. The amplified fragments are used in the
CC   detection of hepatitis virus. This target DNA was isolated from serum
CC   of chronically infected NANBH patients who were C100 antibody-positive
CC   and HCV RNA (NS5 region) positive. Reverse transcription-PCR and PCR
CC   were performed on cDNA and the total human NANBH DNA was constructed
CC   from 23 clones.
XX
SQ   Sequence 9436 BP; 1876 A; 2840 C; 2655 G; 1974 T; 91 other;

Alignment Scores:
Pred. No.:              663          Length:          9436
Score:                  43.00        Matches:         8
Percent Similarity:     100.00%      Conservative:    1
Best Local Similarity:  88.89%       Mismatches:      0
Query Match:            97.73%       Indels:          0
DB:                     15           Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AAQ63499 (1-9436)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||||||||||||||||||||||||| |||
Db       3456 CTACTTGGTTGCATTGTTACTAGCCTC 3482

RESULT 2
AAQ58817
ID   AAQ58817 standard; cDNA; 819 BP.
XX
AC   AAQ58817;
XX
DT   24-NOV-1994  (first entry)
XX
DE   NANBH virus gene fragment #4.
XX
KW   Antigen; structural; non-structural; non A non B hepatitis virus;
KW   NANBHV; NANBH; patient; plasma; diagnosis; detection; carrier; ss.
XX
OS   Non A Non B hepatitis virus.
XX
PN   JP06070778-A.
XX
PD   15-MAR-1994.
XX
PF   01-JUN-1993;   93JP-0156087.
XX
PR   10-JUL-1992;   92JP-0207391.
XX
PA   (KOKU-) KOKUSAI SHIYAKU KK.
PA   (SANW ) SANWA KAGAKU KENKYUSHO CO.
PA   (TOFU ) TONEN CORP.
PA   (TOKR-) ZH TOKYOTO RINSHO IGAKU SOGO KENKYUSHO.
XX
DR   WPI; 1994-128677/16.
DR   P-PSDB; AAR50072.
XX
PT   Nucleic acid fragment coding non-A non-B hepatitis virus antigen
PT   - useful in diagnosis of NANB patient and detection of virus
PT   carrier
XX
PS   Claim 8; Page 18-19; 37pp; Japanese.
XX
CC   The sequences given in AAQ58814-27 encode antigens of structural and
CC   non-structural regions of non A non B hepatitis virus (NANBHV).
CC   These sequences were derived from the plasma of a NANBH patient by
CC   recombinant DNA techniques. These fragments are useful for the
CC   diagnosis of NANBH patients and the detection of NANBHV carriers.
XX
SQ   Sequence 819 BP; 150 A; 255 C; 241 G; 173 T; 0 other;

Alignment Scores:
Pred. No.:              66.4         Length:          819
Score:                  42.00        Matches:         8
Percent Similarity:     100.00%      Conservative:    1
Best Local Similarity:  88.89%       Mismatches:      0
Query Match:            95.45%       Indels:          0
DB:                     15           Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AAQ58817 (1-819)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||||||||||||||||||||||||| |||
Db         67 CTAATTGGCTGCATCATCACCAGCCTC 93

RESULT 4
AAQ64068
ID   AAQ64068 standard; cDNA; 3461 BP.
XX
AC   AAQ64068;
XX
DT   14-FEB-1995  (first entry)
XX
DE   Non-A, non-B hepatitis virus gene #4.
XX
KW   Non-A, non-B hepatitis virus; NANBHV; hepatitis C virus; HCV;
KW   core; ENV; NS1; NS2; NS3; antigen; detection; ss.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             307..3461
FT                   /*tag= a
FT   misc_RNA        307..879
FT                   /*tag= b
FT                   /label= core
FT   misc_RNA        880..1455
FT                   /*tag= c
FT                   /label= ENV
FT   misc_RNA        1456..2736
FT                   /*tag= d
FT                   /label= NS1
FT   misc_RNA        2737..3461
FT                   /*tag= e
FT                   /label= NS2-3
FT                   /note= "NS3 N-terminal"
XX
PN   JP06141870-A.
XX
PD   24-MAY-1994.
XX
PF   12-MAR-1992;   92JP-0088140.
XX
PR   12-MAR-1992;   92JP-0088140.
XX
PA   (SANW ) SANWA KAGAKU KENKYUSHO CO.
```

```
PA   (TOFU ) TONEN CORP.
PA   (TOKR-) ZH TOKYOTO RINSHO IGAKU SOGO KENKYUSHO.
XX
DR   WPI; 1994-205026/25.
DR   P-PSDB; AAR54066.
XX
PT   DNA coding a Non-A, non-B hepatitis virus antigen - useful for
PT   detecting HCV within serum
XX
PS   Claim 1-5; Page 11-15; 22pp; Japanese.
XX
CC   Hepatitis C virus #4 and #6 genes were isolated (AAQ64068-69).
CC   Both genes contain the core, ENV, NS1, NS2 and NS3 regions.
CC   A core region fragment is given in AAQ64067.
XX
SQ   Sequence 3461 BP; 638 A; 1046 C; 1012 G; 765 T; 0 other;

Alignment Scores:
Pred. No.:                336           Length:          3461
Score:                    42.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              95.45%        Indels:          0
DB:                       15            Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AAQ64068 (1-3461)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |.|:::|||||||||||||||||||||
Db       3421 CTAATTGGCTGCATCATCACCAGCCTA 3447

RESULT 4
AAT30386
ID   AAT30386 standard; cDNA; 3461 BP.
XX
AC   AAT30386;
XX
DT   22-AUG-1996  (first entry)
XX
DE   5'UTR/CORE/ENV/NS1/NS2/NS3 cDNA from HCV (#4).
XX
KW   Hepatitis C virus; HCV; antigen; detection; antibody; ds.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             307..3461
FT                   /*tag= a
FT   misc_feature    307..879
FT                   /*tag= b
FT                   /product= Core peptide
FT   misc_feature    880..1455
FT                   /*tag= c
FT                   /product= ENV1
FT   misc_feature    1456..2736
FT                   /*tag= d
FT                   /product= NS1/ENV2
FT   misc_feature    2737..3461
FT                   /*tag= e
FT                   /product= NS2 and NS3
XX
PN   JP07133291-A.
XX
PD   23-MAY-1995.
XX
PF   18-JUN-1993;   93JP-0147944.
XX
PR   18-JUN-1993;   93JP-0147944.
XX
PA   (TOFU ) TONEN CORP.
XX
DR   WPI; 1995-220780/29.
DR   P-PSDB; AAR98361.
XX
PT   Recombinant polypeptide comprising partial NS1 region of hepatitis
PT   non A non-B viral antigen - used in a method for detecting
PT   antibodies against hepatitis non-A non-B virus.
XX
PS   Disclosure; Page 10-12; 15pp; Japanese.
XX
CC   The sequences given in AAT30386-87 encode the 5'UTR/CORE/ENV/NS1/NS2/
CC   NS3 protein region derived from hepatitis C virus (HCV) isolates #4
CC   and #6 respectively.  The proteins encoded by these sequences partic.
CC   encode amino acids 384-495 of the HCV NS1 antigen.  These protein
CC   fragments may be used in the detection of antibodies against HCV.
XX
SQ   Sequence 3461 BP; 638 A; 1046 C; 1012 G; 765 T; 0 other;

Alignment Scores:
Pred. No.:                336           Length:          3461
Score:                    42.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              95.45%        Indels:          0
DB:                       16            Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AAT30386 (1-3461)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||:::|||||.|||||||||||||||
Db       3421 CTAATTGGCTGCATCATCACCAGCCTA 3447

RESULT 5
AAS86668/c
ID   AAS86668 standard; cDNA; 4977 BP.
XX
AC   AAS86668;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #22472.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   25-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG22481.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 22472; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
```

CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4977 BP; 1417 A; 1392 C; 1202 G; 966 T; 0 other;

Alignment Scores:

Pred. No.:	795	Length:	4977
Score:	41.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	77.78%	Mismatches:	0
Query Match:	93.18%	Indels:	0
DB:	23	Gaps:	0

US-09-965-594-1_COPY_14_22 (1-9) x AAS86668 (1-4977)

Qy 1 LeuLeuGlyCysIleIleThrSerLeu 9
Db 1530 CTTTTGGGGTGTGTTTTGACTTCTCTG 1504

RESULT 6
AAS88891/c
ID AAS88891 standard; cDNA; 4977 BP.
XX
AC AAS88891;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #24695.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG24704.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 24695; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4977 BP; 1417 A; 1392 C; 1202 G; 966 T; 0 other;

Alignment Scores:

Pred. No.:	795	Length:	4977
Score:	41.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	77.78%	Mismatches:	0
Query Match:	93.18%	Indels:	0
DB:	23	Gaps:	0

US-09-965-594-1_COPY_14_22 (1-9) x AAS88891 (1-4977)

Qy 1 LeuLeuGlyCysIleIleThrSerLeu 9
Db 1530 CTTTTGGGGTGTGTTTTGACTTCTCTG 1504

RESULT 7
AAS89842/c
ID AAS89842 standard; cDNA; 4977 BP.
XX
AC AAS89842;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #25646.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG25655.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 25646; 103pp; English.
XX

```
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4977 BP; 1417 A; 1392 C; 1202 G; 966 T; 0 other;

Alignment Scores:
Pred. No.:              795          Length:          4977
Score:                  41.00        Matches:         7
Percent Similarity:     100.00%      Conservative:    2
Best Local Similarity:  77.78%       Mismatches:      0
Query Match:            93.18%       Indels:          0
DB:                     23           Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AAS89842 (1-4977)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||||||||||||.:::::: |||||||
Db       1530 CTTTTGGGGTGIGTTTTGACTTCTCTG 1504

RESULT 9
AAX24843
ID   AAX24843 standard; DNA; 9595 BP.
XX
AC   AAX24843;
XX
DT   21-JUN-1999  (first entry)
XX
DE   Infectious hepatitis C virus genotype 1b strain HC-J4 genome.
XX
KW   HCV; infectious clone; infection; diagnosis; therapy; vaccine;
KW   screening; assay; antiviral; virucide; ss.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             342..9374
FT                   /*tag= a
XX
PN   WO9904008-A2.
XX
PD   28-JAN-1999.
XX
PF   16-JUL-1998;    98WO-US14688.
XX
PR   27-JAN-1998;    98US-0014416.
PR   18-JUL-1997;    97US-0053062.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Bukh J,  Emerson SU,  Purcell RH,  Yanagi M;
XX
DR   WPI; 1999-132252/11.
DR   P-PSDB; AAW98022.
XX
PT   New isolated hepatitis C virus nucleic acids - used to develop
PT   products for the diagnosis, prevention and treatment of HCV
PT   infections and for developing screening assays
XX
PS   Claim 3; Fig 14A-F; 126pp; English.
XX
CC   The present sequence comprises the nucleic acid sequence of the
CC   genome of infectious hepatitis C virus (HCV) genotype 1b strain
CC   HC-J4 (ATCC 209596) that is capable of expressing this virus when
CC   transfected into cells.  HC-J4 was obtained from acute phase plasma
CC   of a chimpanzee experimentally infected with serum containing
CC   HC-J4/91.  The claimed infectious nucleic acid sequence can be used
CC   to produce chimeric genomes (see AAX24833) consisting of the open
CC   reading frames of infectious nucleic acid sequences of other
CC   genotypes (including genotypes 1-6) and subtypes (such as 1b, 2a,
CC   2b, 2c, 3a, 4a-f, 5a and 6a) of HCV.  The invention also relates to
CC   the introduction of mutations or deletions into infectious nucleic
CC   acid sequences in order to produce an attenuated HCV virus suitable
CC   for vaccine development.  Infectious nucleic acid sequences can
CC   also be used to produce attenuated virus via passage in vitro or in
CC   vivo of the viruses produced by transfection of a host cell with
CC   the infectious nucleic acid sequence.  Vaccines comprising one or
CC   more polypeptides made from the infectious nucleic acid sequence are
CC   used to immunise mammals, especially humans, against hepatitis C.
CC   The nucleic acid sequences can also be used to induce protective
CC   immunity against the virus.  The nucleic acid sequences or their
CC   encoded proteases (e.g. NS3 protease) can additionally be used to
CC   develop screening assays to identify antiviral agents for HCV.
XX
SQ   Sequence 9595 BP; 1934 A; 2842 C; 2698 G; 2121 T; 0 other;

Alignment Scores:
Pred. No.:              1.66e+03     Length:          9595
Score:                  41.00        Matches:         8
Percent Similarity:     100.00%      Conservative:    1
Best Local Similarity:  88.89%       Mismatches:      0
Query Match:            93.18%       Indels:          0
DB:                     20           Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AAX24843 (1-9595)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              :::.|||||||||||||||||||||||
Db       3456 GTACTTGGTTGCATCATCACTAGCCTC 3482

RESULT 9
AAC86939
ID   AAC86939 standard; DNA; 9595 BP.
XX
AC   AAC86939;
XX
DT   02-APR-2001  (first entry)
XX
DE   Nucleotide sequence of a hepatitis C virus (HCV) clone genotype 1b.
XX
KW   Chimeric virus; bovine viral diarrhoea virus; BVDV; hepatitis C virus;
KW   HCV; vaccine; viral inhibitor; antiviral; ss.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             342..9374
FT                   /*tag= a
XX
PN   WO200075352-A2.
XX
PD   14-DEC-2000.
XX
PF   02-JUN-2000; 2000WO-US15527.
XX
PR   04-JUN-1999;    99US-0137817.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
```

XX
PI Nam J, Bukh J, Emerson SU, Purcell RH;
XX
DR WPI; 2001-071081/08.
DR P-PSDB; AAB31170.
XX
PT New nucleic acid comprising a chimeric bovine viral diarrhoea virus
PT genome in which the (non-)structural region has been replaced by
PT hepatitis C virus (HCV) genome useful for treating or preventing HCV
PT signs and symptoms -
XX
PS Disclosure; Fig 4A-F; 97pp; English.
XX
CC The specification describes a nucleic acid comprising a chimeric virus
CC genome, specifically bovine viral diarrhoea virus (BVDV) genome in which
CC the (non-)structural region has been replaced by the (non-)structural
CC region of a hepatitis C virus (HCV) genome. The nucleic acids comprising
CC the chimeric virus and the chimeric virus are useful for identifying
CC cell lines capable of supporting the replication of these chimeric
CC viruses, in screening for neutralizing antibodies to HCV of different
CC genotypes, in the production of HCV-BVDV virions, for the development
CC of inactivated or attenuated vaccines to prevent HCV-BVDV in a mammal,
CC in studying the molecular properties of HCV indirectly in vitro, and in
CC identifying inhibitors of viral enzyme activity which would be useful
CC as antiviral agents. Formulations or compositions comprising the
CC chimeric virions may be used to treat or prevent the signs and symptoms
CC of HCV. The present sequence represents a HCV clone, which is used
CC to construct chimeric nucleic acids of the invention.
XX
SQ Sequence 9595 BP; 1934 A; 2842 C; 2698 G; 2121 T; 0 other;

Alignment Scores:

Pred. No.:	1.66e-03	Length:	9595
Score:	41.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	93.18%	Indels:	0
DB:	22	Gaps:	0

US-09-965-594-1_COPY_14_22 (1-9) x AAC86939 (1-9595)

Qy 1 LeuLeuGlyCysIleIleThrSerLeu 9
:::|||:|||||:|| |||||:|||||
Db 3456 GTACTTGGTTGCATCATCACTAGCCTC 3482

RESULT 10
AAF23492
ID AAF23492 standard; DNA; 9595 BP.
XX
AC AAF23492;
XX
DT 21-MAR-2001 (first entry)
XX
DE Infectious Hepatitis C virus 1b genotype.
XX
KW GBV-B; hepatitis C virus; HCV; vaccine; ds.
XX
OS Hepatitis C virus.
XX
PN WO200075337-A1.
XX
PD 14-DEC-2000.
XX
PF 02-JUN-2000; 2000WO-US15293.
XX
PR 04-JUN-1999; 99US-0137694.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Bukh J, Yanagi M, Emerson SU, Purcell RH;
XX
DR WPI; 2001-091214/10.
XX
PT New infectious nucleic acids of the GB virus-B clone, useful for
PT indirectly studying the molecular properties of hepatitis C virus (HCV)
PT and in developing vaccines and therapeutics for HCV -
XX
PS Disclosure; Fig 7; 96pp; English.
XX
CC The present invention relates to GB virus-B. The nucleic acid molecules
CC of the invention are useful for indirectly studying the molecular
CC properties of hepatitis C virus (HCV). The infectious nucleic acid
CC sequence of the GB virus-B clone and the HCV/GBV-B chimeras may be used
CC in the development of vaccines and therapeutics for HCV.
XX
SQ Sequence 9595 BP; 1934 A; 2843 C; 2697 G; 2121 T; 0 other;

Alignment Scores:

Pred. No.:	1.66e+03	Length:	9595
Score:	41.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	93.18%	Indels:	0
DB:	22	Gaps:	0

US-09-965-594-1_COPY_14_22 (1-9) x AAF23492 (1-9595)

Qy 1 LeuLeuGlyCysIleIleThrSerLeu 9
:::|||:|||||:|| |||||:|||||
Db 3456 GTACTTGGTTGCATCATCACTAGCCTC 3482

RESULT 11
AAF22302
ID AAF22302 standard; DNA; 94895 BP.
XX
AC AAF22302;
XX
DT 20-MAR-2001 (first entry)
XX
DE BAC containing repeats from centromeres 1-4 #25.
XX
KW Centromere; michrosome; vector; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200055325-A2.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-US07392.
XX
PR 18-MAR-1999; 99US-0125219.
PR 01-APR-1999; 99US-0127409.
PR 18-MAY-1999; 99US-0134770.
PR 13-SEP-1999; 99US-0153584.
PR 17-SEP-1999; 99US-0154603.
XX
PA (UYCH-) UNIV CHICAGO.
XX
PI Preuss D, Copenhaver G, Keith K;
XX
DR WPI; 2000-587529/55.
XX
PT Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited michrosomes which can serve as vectors for
PT the construction of transgenic plant and animal cells -
XX
PS Claim 102; Page 804-819; 1449pp; English.
XX
CC The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited michrosomes which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
XX

```
SQ   Sequence 94895 BP; 28943 A; 18101 C; 18466 G; 29385 T; 0 other;

Alignment Scores:
Pred. No.:              2.2e+04        Length:         94895
Score:                  41.00          Matches:        7
Percent Similarity:     100.00%        Conservative:   2
Best Local Similarity:  77.78%         Mismatches:     0
Query Match:            93.18%         Indels:         0
DB:                     21             Gaps:           0

US-09-965-594-1_COPY_14_22 (1-9) x AAF22302 (1-94895)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||||||||||||.|||:::|||||:::
Db       3672 TTGTTAGGTTGCATTGTGACGAGTATG 3698

RESULT 12
AAF22302/c
ID   AAF22302 standard; DNA; 94895 BP.
XX
AC   AAF22302;
XX
DT   20-MAR-2001  (first entry)
XX
DE   BAC containing repeats from centromeres 1-4 #25.
XX
KW   Centromere; michrosome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US07392.
XX
PR   18-MAR-1999;   99US-0125219.
PR   01-APR-1999;   99US-0127409.
PR   18-MAY-1999;   99US-0134770.
PR   13-SEP-1999;   99US-0153584.
PR   17-SEP-1999;   99US-0154603.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D,  Copenhaver G,  Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited michrosomes which can serve as vectors for
PT   the construction of transgenic plant and animal cells  -
XX
PS   Claim 102; Page 804-819; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited michrosomes which can serve as vectors for
CC   the construction of transgenic plant and animal cells expressing
CC   selected proteins such as hormones, enzymes, interleukins, clotting
CC   factors, cytokines, antibodies, and growth factors.
XX
SQ   Sequence 94895 BP; 28943 A; 18101 C; 18466 G; 29385 T; 0 other;

Alignment Scores:
Pred. No.:              2.2e-04        Length:         94895
Score:                  41.00          Matches:        7
Percent Similarity:     100.00%        Conservative:   2
Best Local Similarity:  77.78%         Mismatches:     0
Query Match:            93.18%         Indels:         0
DB:                     21             Gaps:           0

US-09-965-594-1_COPY_14_22 (1-9) x AAF22302 (1-94895)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||||||||||||.|||:::|||||:::
Db       44375 TTGTTAGGTTGCATTGTGACGAGTATG 44349

RESULT 13
AAF22287
ID   AAF22287 standard; DNA; 103929 BP.
XX
AC   AAF22287;
XX
DT   20-MAR-2001  (first entry)
XX
DE   BAC containing repeats from centromeres 1-4 #10.
XX
KW   Centromere; michrosome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US07392.
XX
PR   18-MAR-1999;   99US-0125219.
PR   01-APR-1999;   99US-0127409.
PR   18-MAY-1999;   99US-0134770.
PR   13-SEP-1999;   99US-0153584.
PR   17-SEP-1999;   99US-0154603.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D,  Copenhaver G,  Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited michrosomes which can serve as vectors for
PT   the construction of transgenic plant and animal cells  -
XX
PS   Claim 102; Page 484-508; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited michrosomes which can serve as vectors for
CC   the construction of transgenic plant and animal cells expressing
CC   selected proteins such as hormones, enzymes, interleukins, clotting
CC   factors, cytokines, antibodies, and growth factors.
XX
SQ   Sequence 103929 BP; 32987 A; 19310 C; 18547 G; 33085 T; 0 other;

Alignment Scores:
Pred. No.:              2.43e-04       Length:         103929
Score:                  41.00          Matches:        7
Percent Similarity:     100.00%        Conservative:   2
Best Local Similarity:  77.78%         Mismatches:     0
Query Match:            93.18%         Indels:         0
DB:                     21             Gaps:           0

US-09-965-594-1_COPY_14_22 (1-9) x AAF22287 (1-103929)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||||||||||||.|||:::|||||:::
Db       5120 TTGTTAGGTTGCATTGTGACGAGTATG 5146

RESULT 14
AAF22286
ID   AAF22286 standard; DNA; 134499 BP.
XX
AC   AAF22286;
XX
DT   20-MAR-2001  (first entry)
XX
```

```
DE   BAC containing repeats from centromeres 1-4 #9.
XX
KW   Centromere; michrosome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US07392.
XX
PR   18-MAR-1999;   99US-0125219.
PR   01-APR-1999;   99US-0127409.
PR   18-MAY-1999;   99US-0134770.
PR   13-SEP-1999;   99US-0153584.
PR   17-SEP-1999;   99US-0154603.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D,  Copenhaver G,  Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited michrosomes which can serve as vectors for
PT   the construction of transgenic plant and animal cells -
XX
PS   Claim 102; Page 453-484; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited michrosomes which can serve as vectors for
CC   the construction of transgenic plant and animal cells expressing
CC   selected proteins such as hormones, enzymes, interleukins, clotting
CC   factors, cytokines, antibodies, and growth factors.
XX
SQ   Sequence 134499 BP; 41565 A; 25130 C; 25225 G; 42577 T; 2 other;

Alignment Scores:
Pred. No.:              3.25e-04        Length:         134499
Score:                  41.00           Matches:        7
Percent Similarity:     100.00%         Conservative:   2
Best Local Similarity:  77.78%          Mismatches:     0
Query Match:            93.18%          Indels:         0
DB:                     21              Gaps:           0

US-09-965-594-1_COPY_14_22 (1-9) x AAF22286 (1-134499)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||||:|||||| ||:::|||||:::
Db       1360 TTGTTAGGTTGCATTGTGACGAGTATG 1386

RESULT 15
AAF22305
ID   AAF22305 standard; DNA; 1082138 BP.
XX
AC   AAF22305;
XX
DT   20-MAR-2001  (first entry)
XX
DE   Arabidopsis thaliana chromosome 4 centromere.
XX
KW   Centromere; michrosome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US07392.
XX
PR   18-MAR-1999;   99US-0125219.
PR   01-APR-1999;   99US-0127409.
PR   18-MAY-1999;   99US-0134770.
PR   13-SEP-1999;   99US-0153584.
PR   17-SEP-1999;   99US-0154603.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D,  Copenhaver G,  Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited michrosomes which can serve as vectors for
PT   the construction of transgenic plant and animal cells -
XX
PS   Claim 68; Page 977-1388; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited michrosomes which can serve as vectors for
CC   the construction of transgenic plant and animal cells expressing
CC   selected proteins such as hormones, enzymes, interleukins, clotting
CC   factors, cytokines, antibodies, and growth factors.
XX
SQ   Sequence 1082138 BP; 348775 A; 194404 C; 195515 G; 343444 T; 0 other;

Alignment Scores:
Pred. No.:              3.29e+05        Length:         1082138
Score:                  41.00           Matches:        7
Percent Similarity:     100.00%         Conservative:   2
Best Local Similarity:  77.78%          Mismatches:     0
Query Match:            93.18%          Indels:         0
DB:                     21              Gaps:           0

US-09-965-594-1_COPY_14_22 (1-9) x AAF22305 (1-1082138)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||||:|||||.|||:::|||||.|:::
Db     885059 TTGTTAGGTTGCATTGTGACGAGTATG 885085


Search completed: September 29, 2003, 19:19:33
Job time : 244 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 29, 2003, 19:11:08 ; Search time 1635 Seconds
(without alignments)
134.629 Million cell updates/sec

Title: US-09-965-594-1_COPY_14_22
Perfect score: 44
Sequence: 1 LLGCIITSL 9

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562778

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 99%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09965594/runat_29092003_164249_14082/app_query.fasta_1.199
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=99.9 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594_@CGN_1_1_2810_@runat_29092003_164249_14082 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	41	93.2	108	28	BH241427	BH241427 AUILF55TR
2	41	93.2	230	29	AL950918	AL950918 Arabidops
c 3	41	93.2	334	29	BZ352875	BZ352875 SALK_0827
c 4	41	93.2	400	29	CNS00Q21	AL085272 Arabidops
5	41	93.2	424	28	BH758372	BH758372 SALK_0185
c 6	41	93.2	444	10	BF651538	BF651538 274341 MA
c 7	41	93.2	465	10	BF651537	BF651537 274340 MA
c 8	41	93.2	498	28	B96162	B96162 F26H1TFB IG
c 9	41	93.2	505	28	B26036	B26036 T5G2TF TAMU
c 10	41	93.2	651	29	BZ289143	BZ289143 SALK_0225
11	41	93.2	683	28	BH241556	BH241556 AUILE77TF
12	41	93.2	693	28	B95974	B95974 F21D4TFC IG
13	41	93.2	719	28	BH241494	BH241494 AUILF22TR
14	41	93.2	724	28	AQ955329	AQ955329 LERAD34TF
15	41	93.2	758	28	BH241165	BH241165 AUILF85TF
c 16	41	93.2	787	28	BH241117	BH241117 AUILB95TF
17	41	93.2	823	28	BH241591	BH241591 AUILG55TF
c 18	41	93.2	831	28	BH241228	BH241228 AUILD16TF
19	41	93.2	835	29	CC429018	CC429018 PUHJV59TD
20	41	93.2	847	29	CC363288	CC363288 PUHLI0TD
21	41	93.2	917	10	BF167496	BF167496 601775095
22	40	90.9	287	9	AI800243	AI800243 ti76d09.x
23	40	90.9	309	9	AI587235	AI587235 tq21f04.x
24	40	90.9	355	14	Z39665	Z39665 HSC1GA052 n
25	40	90.9	384	9	AI583735	AI583735 tt03h01.x
26	40	90.9	401	14	CA902724	CA902724 PCS01861X
27	40	90.9	406	10	BF109646	BF109646 7166c06.x
28	40	90.9	407	10	BF727441	BF727441 by21e11.x
29	40	90.9	413	28	AQ759141	AQ759141 HS_2201_A
30	40	90.9	423	9	AA721993	AA721993 zh17d10.s
31	40	90.9	424	14	T03709	T03709 IB779 Infan
32	40	90.9	425	9	AA812004	AA812004 ob39e08.s
33	40	90.9	426	9	AW665172	AW665172 hi87q11.x
34	40	90.9	473	9	AI927322	AI927322 wo89e11.x
35	40	90.9	540	12	BQ017316	BQ017316 UI-H-DT1-
c 36	40	90.9	588	10	BE785099	BE785099 601478365
37	40	90.9	588	14	H15452	H15452 ym29g06.s1
38	40	90.9	613	14	H15273	H15273 ym30e09.s1
39	40	90.9	625	12	BM682281	BM682281 UI-E-EJ0-
40	40	90.9	683	29	BZ325209	BZ325209 ic50c11.g
41	40	90.9	702	13	BQ184273	BQ184273 UI-E-EJ1-
c 42	40	90.9	710	13	BQ187309	BQ187309 UI-E-EJ1-
c 43	40	90.9	915	10	BG685367	BG685367 602637211
c 44	40	90.9	927	13	BU527696	BU527696 AGENCOURT
c 45	40	90.9	1006	10	BG758345	BG758345 602712536

ALIGNMENTS

RESULT 1
BH241427/c
LOCUS BH241427 108 bp DNA linear GSS 13-NOV-2001
DEFINITION AUILF55TR AUIL Arabidopsis thaliana genomic clone AUILF55, genomic survey sequence.
ACCESSION BH241427
VERSION BH241427.1 GI:16915855
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

```
REFERENCE   1  (bases 1 to 108)
  AUTHORS   Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
  TITLE     Survey sequencing of Arabidopsis thaliana BAC F26J21
  JOURNAL   Unpublished
COMMENT     Other_GSSs: AUILF55TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 720.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .108
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="AUILF55"
                     /clone_lib="AUIL"
                     /note="Vector: pHOS2; Site_1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
BASE COUNT       36 a      29 c       8 g      35 t
ORIGIN

Alignment Scores:
Pred. No.:                90.9          Length:          108
Score:                    41.00         Matches:         7
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    77.78%        Mismatches:      0
Query Match:              93.18%        Indels:          0
DB:                       28            Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x BH241427 (1-108)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||||||||||||||:::||||||:::
Db         56 TTGTTAGGTTGCATTGTGACGAGTATG 30

RESULT 2
AL950918
LOCUS       AL950918                 230 bp    DNA     linear   GSS 24-OCT-2002
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence GK-332C10-016064,
            genomic survey sequence.
ACCESSION   AL950918
VERSION     AL950918.1  GI:24407540
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H.
            and Weisshaar,B.
  TITLE     A pipeline for automated high-throughput generation of FSTs
            (flanking sequence tags) from Arabidopsis thaliana T-DNA
            transformed lines
  JOURNAL   Unpublished
REFERENCE   2
  AUTHORS   Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.
  TITLE     A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
            for flanking sequence tag based reverse genetics
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 230)
  AUTHORS   Li,Y., Rosso,M., Strizhov,N. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2002) Weisshaar B., Max-Planck-Institut fuer
            Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT     This sequence is recovered from the left border of the T-DNA. It
            indicates an insertion within the locus defined by clone F4M19. The
            sequences are generated at the MPI for Plant Breeding Research in
            the context of the GABI-Kat project. GABI-Kat is part of the German
            Plant Genomics program designated 'GABI'. Information on line
            availability can be found at:
            http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES             Location/Qualifiers
     source          1. .230
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="GK-332C10-016064"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /note="PCR was performed on DNA from Arabidopsis thaliana
                     plants (T1) which were transformed with the T-DNA from
                     vector pAC161. The lines contain one or more T-DNA
                     insertions. The DNA fragment(s) resulting from the PCR
                     were directly sequenced to determine the genomic sequence
                     flanking the insertion. Sequences displaying significant
                     similarity to the A. thaliana nuclear genome sequence were
                     processed for submission. T-DNA derived sequences were
                     removed"
BASE COUNT       67 a      42 c      48 g      73 t
ORIGIN

Alignment Scores:
Pred. No.:                277           Length:          230
Score:                    41.00         Matches:         7
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    77.78%        Mismatches:      0
Query Match:              93.18%        Indels:          0
DB:                       29            Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x AL950918 (1-230)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||||||||||||||:::||||||:::
Db        164 TTGTTAGGTTGCATTGTGACGAGTATG 190

RESULT 3
BZ352875/c
LOCUS       BZ352875                 334 bp    DNA     linear   GSS 14-NOV-2002
DEFINITION  SALK_082734.31.45.x Arabidopsis thaliana TDNA insertion lines
            Arabidopsis thaliana genomic clone SALK_082734.31.45.x, genomic
            survey sequence.
ACCESSION   BZ352875
VERSION     BZ352875.1  GI:24943737
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 334)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
            , Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .334
```

```
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_082734.31.45.x"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains one or more TDNA insertion
                     elements. The resultant fragment for each line was
                     directly sequenced to determine the genomic sequence at
                     the site of insertion. Details of the protocols used can
                     be found at http://signal.salk.edu/tdna_protocols.html"
BASE COUNT      111 a     71 c     60 g     91 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:                480           Length:         334
Score:                    41.00         Matches:        7
Percent Similarity:       100.00%       Conservative:   2
Best Local Similarity:    77.78%        Mismatches:     0
Query Match:              93.18%        Indels:         0
DB:                       29            Gaps:           0

US-09-965-594-1_COPY_14_22 (1-9) x BZ352875 (1-334)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||||||:||||||:|:::|||||||:::
Db        188 TTGTTAGGCTGCATTGTGACGAGTATG 162

RESULT 4
CNS00Q2I/c
LOCUS       CNS00Q2I                 400 bp    DNA     linear   GSS 28-JUN-1999
DEFINITION  Arabidopsis thaliana genome survey sequence T7 end of BAC F10H2 of
            IGF library from strain Columbia of Arabidopsis thaliana, genomic
            survey sequence.
ACCESSION   AL085272
VERSION     AL085272.1  GI:5286412
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Salanoubat,M., Choisne,N., Artiguenave,F., Brottier,P., Wincker,P.,
            Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 400)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .400
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="F10H2"
                     /clone_lib="IGF"
                     /note="end : T7"
BASE COUNT      148 a     84 c     61 g     107 t
ORIGIN

Alignment Scores:
Pred. No.:                627           Length:         400
Score:                    41.00         Matches:        7
Percent Similarity:       100.00%       Conservative:   2
Best Local Similarity:    77.78%        Mismatches:     0
Query Match:              93.18%        Indels:         0
DB:                       29            Gaps:           0

US-09-965-594-1_COPY_14_22 (1-9) x CNS00Q2I (1-400)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||||||:||||||:|:::|||||||:::
Db        350 TTGTTAGGTTGCATTGTGACGAGTATG 324

RESULT 5
BH758372
LOCUS       BH758372                 424 bp    DNA     linear   GSS 01-MAR-2002
DEFINITION  SALK_018581.33.10.x Arabidopsis thaliana TDNA insertion lines
            Arabidopsis thaliana genomic clone SALK_018581.33.10.x, genomic
            survey sequence.
ACCESSION   BH758372
VERSION     BH758372.1  GI:19043622
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
            , Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .424
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_018581.33.10.x"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains one or more TDNA insertion
                     elements. The resultant fragment for each line was
                     directly sequenced to determine the genomic sequence at
                     the site of insertion. Details of the protocols used can
                     be found at http://signal.salk.edu/tdna_protocols.html"
BASE COUNT       62 a    151 c     98 g     113 t
ORIGIN

Alignment Scores:
Pred. No.:                683           Length:         424
Score:                    41.00         Matches:        7
Percent Similarity:       100.00%       Conservative:   2
Best Local Similarity:    77.78%        Mismatches:     0
Query Match:              93.18%        Indels:         0
DB:                       28            Gaps:           0

US-09-965-594-1_COPY_14_22 (1-9) x BH758372 (1-424)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||||||||||:||||||:::|||||||:::
Db         74 TTGTTAGGCTGCATTGTGACGAGTATG 100

RESULT 6
BF651538/c
LOCUS       BF651538                 444 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  274341 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF651538
```

```
VERSION     BF651538.1  GI:11916668
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 444)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
   PUBMED   11282978
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 63  row: B  column: 22
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .444
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 3BOV"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT      116 a     96 c    115 g    117 t
ORIGIN

Alignment Scores:
Pred. No.:                731           Length:         444
Score:                    41.00         Matches:        8
Percent Similarity:       100.00%       Conservative:   1
Best Local Similarity:    88.89%        Mismatches:     0
Query Match:              93.18%        Indels:         0
DB:                       10            Gaps:           0

US-09-965-594-1_COPY_14_22 (1-9) x BF651538 (1-444)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||:::|||||||||||||||||||||
Db        113 TTGGTGGGTTGTATAATTACTTCATTA 87

RESULT 7
BF651537/c
LOCUS       BF651537                 465 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  274340 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF651537
VERSION     BF651537.1  GI:11916667
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 465)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
   PUBMED   11282978
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 63  row: B  column: 21
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .465
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 3BOV"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT      121 a     97 c    122 g    125 t
ORIGIN

Alignment Scores:
Pred. No.:                782           Length:         465
Score:                    41.00         Matches:        8
Percent Similarity:       100.00%       Conservative:   1
Best Local Similarity:    88.89%        Mismatches:     0
Query Match:              93.18%        Indels:         0
DB:                       10            Gaps:           0

US-09-965-594-1_COPY_14_22 (1-9) x BF651537 (1-465)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||:::|||||||||||||||||||||
Db        113 TTGGTGGGTTGTATAATTACTTCATTA 87

RESULT 8
B96162/c
LOCUS       B96162                   498 bp    DNA     linear   GSS 31-MAR-1998
DEFINITION  F26H1TFB IGF Arabidopsis thaliana genomic clone F26H1, genomic
            survey sequence.
ACCESSION   B96162
VERSION     B96162.1  GI:2998638
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K.
            , Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter
            ,J.C.
  TITLE     A BAC End Sequence Database for Identifying Minimal Overlaps in
            Arabidopsis Genomic Sequencing. Update 3
```

```
  JOURNAL   Unpublished
COMMENT     Other_GSSs: F26H1TR
            Contact: Steve Rounsley
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: rounsley@tigr.org
            Seq primer: M13-21
            Class: BAC ends
            High quality sequence stop: 498.
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="F26H1"
                     /sex="hermaphrodite"
                     /clone_lib="IGF"
                     /note="Vector: BeloBACII; Site_1: EcoRI; Site_2: EcoRI;
                     Produced by Thomas Altmann"
BASE COUNT      184 a    105 c     80 g    129 t
ORIGIN

Alignment Scores:
Pred. No.:                866             Length:          498
Score:                    41.00           Matches:         7
Percent Similarity:       100.00%         Conservative:    2
Best Local Similarity:    77.78%          Mismatches:      0
Query Match:              93.18%          Indels:          0
DB:                       28              Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x B96162 (1-498)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||||||||:|||||:::|||||.:::
Db        370 TTGTTAGGTTGCATTGTGACGAGTATG 344

RESULT 9
B26036/c
LOCUS       B26036                   505 bp    DNA     linear   GSS 13-OCT-1997
DEFINITION  T5G2TF TAMU Arabidopsis thaliana genomic clone T5G2, genomic survey
            sequence.
ACCESSION   B26036
VERSION     B26036.1  GI:2512002
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 505)
  AUTHORS   Rounsley,S.D., Kelley,J.M., Field,C.E., Craven,M.B., Adams,M.D. and
            Venter,J.C.
  TITLE     Use of a BAC End Sequence Database To Identify Minimal Overlaps for
            Arabidopsis Genomic Sequencing
  JOURNAL   Unpublished
COMMENT     Other_GSSs: T5G2TR
            Contact: Steve Rounsley
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: rounsley@tigr.org
            Seq primer: M13-21
            Class: BAC ends
            High quality sequence stop: 505.
FEATURES             Location/Qualifiers
     source          1. .505
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="T5G2"
                     /sex="hermaphrodite"
                     /clone_lib="TAMU"
                     /note="Vector: BeloBACII; Site_1: HindIII; Site_2: HindIII
                     ; Produced by Rod Wing"
BASE COUNT      173 a    107 c     93 g    131 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:                884             Length:          505
Score:                    41.00           Matches:         7
Percent Similarity:       100.00%         Conservative:    2
Best Local Similarity:    77.78%          Mismatches:      0
Query Match:              93.18%          Indels:          0
DB:                       28              Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x B26036 (1-505)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||||||||||||||:::|||||:::
Db        196 TTGTTAGGTTGCATTGTGACGAGTATG 170

RESULT 10
BZ289143/c
LOCUS       BZ289143                 651 bp    DNA     linear   GSS 24-OCT-2002
DEFINITION  SALK_022540.51.60.x Arabidopsis thaliana TDNA insertion lines
            Arabidopsis thaliana genomic clone SALK_022540.51.60.x, genomic
            survey sequence.
ACCESSION   BZ289143
VERSION     BZ289143.1  GI:24330427
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 651)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
            , Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .651
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_022540.51.60.x"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains one or more TDNA insertion
                     elements. The resultant fragment for each line was
                     directly sequenced to determine the genomic sequence at
                     the site of insertion. Details of the protocols used can
                     be found at http://signal.salk.edu/tdna_protocols.html"
BASE COUNT      207 a    131 c    113 g    197 t      3 others
ORIGIN
```

```
Alignment Scores:
Pred. No.:                1.29e+03        Length:          651
Score:                    41.00           Matches:         7
Percent Similarity:       100.00%         Conservative:    2
Best Local Similarity:    77.78%          Mismatches:      0
Query Match:              93.18%          Indels:          0
DB:                       29              Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x BZ289143 (1-651)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||||||||||||||:::|||||||:::
Db        170 TTGTTAGGTTGCATTGTGACGAGTATG 144

RESULT 11
BH241556
LOCUS       BH241556                683 bp    DNA     linear   GSS 13-NOV-2001
DEFINITION  AUILE77TF AUIL Arabidopsis thaliana genomic clone AUILE77, genomic
            survey sequence.
ACCESSION   BH241556
VERSION     BH241556.1  GI:16916061
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 683)
  AUTHORS   Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
  TITLE     Survey sequencing of Arabidopsis thaliana BAC F26J21
  JOURNAL   Unpublished
COMMENT     Other_GSSs: AUILE77TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3524
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 720.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .683
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="AUILE77"
                     /clone_lib="AUIL"
                     /note="Vector: pHOS2; Site_1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
BASE COUNT      196 a    129 c    138 g    220 t
ORIGIN

Alignment Scores:
Pred. No.:                1.38e-03        Length:          683
Score:                    41.00           Matches:         7
Percent Similarity:       100.00%         Conservative:    2
Best Local Similarity:    77.78%          Mismatches:      0
Query Match:              93.18%          Indels:          0
DB:                       28              Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x BH241556 (1-683)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||||||||||||||:::|||||||:::
Db        655 TTGTTAGGTTGCATTGTGACGAGTATG 681

RESULT 12
B95974
LOCUS       B95974                  693 bp    DNA     linear   GSS 31-MAR-1998
DEFINITION  F21D4TFC IGF Arabidopsis thaliana genomic clone F21D4, genomic
            survey sequence.
ACCESSION   B95974
VERSION     B95974.1  GI:2998450
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 693)
  AUTHORS   Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and
            Venter,J.C.
  TITLE     A BAC End Sequence Database for Identifying Minimal Overlaps in
            Arabidopsis Genomic Sequencing. Update 3
  JOURNAL   Unpublished
COMMENT     Contact: Steve Rounsley
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: rounsley@tigr.org
            Seq primer: M13-21
            Class: BAC ends
            High quality sequence stop: 693.
FEATURES             Location/Qualifiers
     source          1. .693
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="F21D4"
                     /sex="hermaphrodite"
                     /clone_lib="IGF"
                     /note="Vector: BeloBACII; Site_1: EcoRI; Site_2: EcoRI;
                     Produced by Thomas Altmann"
BASE COUNT      210 a    117 c    144 g    221 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:                1.41e+03        Length:          693
Score:                    41.00           Matches:         7
Percent Similarity:       100.00%         Conservative:    2
Best Local Similarity:    77.78%          Mismatches:      0
Query Match:              93.18%          Indels:          0
DB:                       28              Gaps:            0

US-09-965-594-1_COPY_14_22 (1-9) x B95974 (1-693)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              |||||||||||||||:::|||||||:::
Db        480 TTGTTAGGTTGCATTGTGACGAGTATG 506

RESULT 13
BH241494
LOCUS       BH241494                719 bp    DNA     linear   GSS 13-NOV-2001
DEFINITION  AUILF22TR AUIL Arabidopsis thaliana genomic clone AUILF22, genomic
            survey sequence.
ACCESSION   BH241494
VERSION     BH241494.1  GI:16915962
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 719)
  AUTHORS   Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
  TITLE     Survey sequencing of Arabidopsis thaliana BAC F26J21
  JOURNAL   Unpublished
COMMENT     Other_GSSs: AUILF22TF
```

```
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 720.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .719
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="AUILF22"
                     /clone_lib="AUIL"
                     /note="Vector: pHOS2; Site_1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
BASE COUNT      167 a    149 c    140 g    233 t
ORIGIN

Alignment Scores:
Pred. No.:              1.49e+03        Length:         719
Score:                  41.00           Matches:        7
Percent Similarity:     100.00%         Conservative:   2
Best Local Similarity:  77.78%          Mismatches:     0
Query Match:            93.18%          Indels:         0
DB:                     28              Gaps:           0

US-09-965-594-1_COPY_14_22 (1-9) x BH241494 (1-719)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||||||||||||||||:::|||||||:::
Db        130 TTGTTAGGTTGCATTGTGACGAGTATG 156

RESULT 14
AQ955329
LOCUS       AQ955329                724 bp    DNA     linear   GSS 28-JAN-2000
DEFINITION  LERAD34TF LERA Arabidopsis thaliana genomic clone LERAD34, genomic
            survey sequence.
ACCESSION   AQ955329
VERSION     AQ955329.1  GI:6783030
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 724)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterback,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished
COMMENT     Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Seq primer: TF
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .724
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /clone="LERAD34"
                     /clone_lib="LERA"
                     /note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was
                     sheared to 0.9-1 Kbp before ligation."
BASE COUNT      187 a    151 c    143 g    243 t
ORIGIN

Alignment Scores:
Pred. No.:              1.5e+03         Length:         724
Score:                  41.00           Matches:        7
Percent Similarity:     100.00%         Conservative:   2
Best Local Similarity:  77.78%          Mismatches:     0
Query Match:            93.18%          Indels:         0
DB:                     28              Gaps:           0

US-09-965-594-1_COPY_14_22 (1-9) x AQ955329 (1-724)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||||||||||||||||:::|||||||:::
Db         70 TTGTTAGGTTGCATTGTGACGAGTATG 96

RESULT 15
BH241165
LOCUS       BH241165                758 bp    DNA     linear   GSS 13-NOV-2001
DEFINITION  AUILF85TF AUIL Arabidopsis thaliana genomic clone AUILF85, genomic
            survey sequence.
ACCESSION   BH241165
VERSION     BH241165.1  GI:16915368
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 758)
  AUTHORS   Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
  TITLE     Survey sequencing of Arabidopsis thaliana BAC F26J21
  JOURNAL   Unpublished
COMMENT     Other_GSSs: AUILF85TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 720.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .758
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="AUILF85"
                     /clone_lib="AUIL"
                     /note="Vector: pHOS2; Site_1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
BASE COUNT      202 a    140 c    157 g    259 t
ORIGIN

Alignment Scores:
Pred. No.:              1.61e+03        Length:         758
Score:                  41.00           Matches:        7
Percent Similarity:     100.00%         Conservative:   2
Best Local Similarity:  77.78%          Mismatches:     0
Query Match:            93.18%          Indels:         0
DB:                     28              Gaps:           0

US-09-965-594-1_COPY_14_22 (1-9) x BH241165 (1-758)

Qy          1 LeuLeuGlyCysIleIleThrSerLeu 9
              ||||| |||||||||:::|||||||:::
Db        318 TTGTTAGGTTGCATTGTGACGAGTATG 344
```

Search completed: September 29, 2003, 20:22:55
Job time : 1632 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 29, 2003, 19:00:07 ; Search time 30 Seconds
(without alignments)
77.416 Million cell updates/sec

Title: US-09-965-594-1_COPY_14_22
Perfect score: 44
Sequence: 1 LLGCIITSL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830296

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 99%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	97.7	180	12	Q8QW08	Q8qw08 hepatitis c
2	43	97.7	180	12	Q8QW28	Q8qw28 hepatitis c
3	43	97.7	180	12	Q8QW27	Q8qw27 hepatitis c
4	43	97.7	180	12	Q8QVX3	Q8qvx3 hepatitis c
5	43	97.7	403	12	Q8QP94	Q8qp94 hepatitis c
6	43	97.7	3010	12	Q9QIY5	Q9qiy5 hepatitis c
7	43	97.7	3010	12	Q9QIY6	Q9qiy6 hepatitis c
8	43	97.7	3010	12	Q9QIY2	Q9qiy2 hepatitis c
9	43	97.7	3010	12	Q8V638	Q8v638 hepatitis c
10	43	97.7	3010	12	Q81541	Q81541 hepatitis c
11	43	97.7	3010	12	Q9DTD6	Q9dtd6 hepatitis c
12	43	97.7	3011	12	Q913D4	Q913d4 hepatitis c
13	42	95.5	181	12	Q91RP5	Q91rp5 hepatitis c
14	42	95.5	3010	12	Q9J3H9	Q9j3h9 hepatitis c
15	42	95.5	3010	12	Q9J3G5	Q9j3g5 hepatitis c
16	41	93.2	180	12	Q8QW19	Q8qw19 hepatitis c
17	41	93.2	361	12	Q70817	Q70817 hepatitis c
18	41	93.2	2864	12	Q92973	Q92973 hepatitis c
19	41	93.2	2864	12	Q9WLK8	Q9wlk8 hepatitis c
20	41	93.2	2864	12	Q92976	Q92976 hepatitis c
21	41	93.2	3010	12	Q02829	Q02829 hepatitis c
22	41	93.2	3010	12	Q92969	Q92969 hepatitis c
23	41	93.2	3010	12	Q02828	Q02828 hepatitis c
24	41	93.2	3010	12	Q9J3G0	Q9j3g0 hepatitis c
25	40	90.9	180	12	Q8QVX2	Q8qvx2 hepatitis c
26	40	90.9	180	12	Q8QW22	Q8qw22 hepatitis c
27	40	90.9	180	12	Q8QW09	Q8qw09 hepatitis c
28	40	90.9	180	12	Q8QW26	Q8qw26 hepatitis c
29	40	90.9	180	12	Q8QVW8	Q8qvw8 hepatitis c
30	40	90.9	180	12	Q8QVX0	Q8qvx0 hepatitis c
31	40	90.9	180	12	Q8QVW2	Q8qvw2 hepatitis c
32	40	90.9	180	12	Q8QVZ2	Q8qvz2 hepatitis c
33	40	90.9	180	12	Q8QVZ6	Q8qvz6 hepatitis c
34	40	90.9	181	12	Q91RQ7	Q91rq7 hepatitis c
35	40	90.9	181	12	Q91RQ6	Q91rq6 hepatitis c
36	40	90.9	2864	12	Q9WLK9	Q9wlk9 hepatitis c
37	40	90.9	2864	12	Q9WLL0	Q9wll0 hepatitis c
38	40	90.9	2864	12	Q92975	Q92975 hepatitis c
39	40	90.9	2864	12	Q92974	Q92974 hepatitis c
40	40	90.9	3010	12	P90192	P90192 hepatitis c
41	40	90.9	3010	12	Q9QIY7	Q9qiy7 hepatitis c
42	40	90.9	3010	12	Q92970	Q92970 hepatitis c
43	40	90.9	3010	12	Q9J3G9	Q9j3g9 hepatitis c
44	40	90.9	3010	12	Q9J3G4	Q9j3g4 hepatitis c
45	40	90.9	3010	12	Q9QIY8	Q9qiy8 hepatitis c

ALIGNMENTS

RESULT 1
Q8QW08
ID Q8QW08 PRELIMINARY; PRT; 180 AA.
AC Q8QW08;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis C virus type 1b.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31647;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L08;
RA Ogata S.;
RT "CORRELATION BETWEEN SECONDARY STRUCTURE OF AN AMINO-TERMINAL PORTION
RT OF THE NONSTRUCTURAL PROTEIN 3 OF HEPATITIS C VIRUS AND DEVELOPMENT OF
RT HEPATOCELLULAR CARCINOMA.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB072067; BAB88249.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
FT NON_TER 1 1
FT NON_TER 180 180
SQ SEQUENCE 180 AA; 18982 MW; 77E64C2673D89EEC CRC64;

Query Match 97.7%; Score 43; DB 12; Length 180;
Best Local Similarity 88.9%; Pred. No. 1.1;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLGCIITSL 9
|:||||:||
Db 13 LLGCIVTSL 21

RESULT 2
Q8QW28
ID Q8QW28 PRELIMINARY; PRT; 180 AA.

```
AC   Q8QW28;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Polyprotein (Fragment).
OS   Hepatitis C virus type 1b.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31647;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H05-4;
RA   Ogata S.;
RT   "CORRELATION BETWEEN SECONDARY STRUCTURE OF AN AMINO-TERMINAL PORTION
RT   OF THE NONSTRUCTURAL PROTEIN 3 OF HEPATITIS C VIRUS AND DEVELOPMENT OF
RT   HEPATOCELLULAR CARCINOMA.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB072047; BAB88229.1; -.
DR   InterPro; IPR004109; HCV_NS3.
DR   Pfam; PF02907; HCV_NS3; 1.
FT   NON_TER       1      1
FT   NON_TER     180    180
SQ   SEQUENCE   180 AA;  18891 MW;  85AF5CE1CF2BCA30 CRC64;

  Query Match             97.7%;  Score 43;  DB 12;  Length 180;
  Best Local Similarity   88.9%;  Pred. No. 1.1;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              |||||:|||
Db         13 LLGCIVTSL 21


RESULT 3
Q8QW27
ID   Q8QW27      PRELIMINARY;      PRT;   180 AA.
AC   Q8QW27;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Polyprotein (Fragment).
OS   Hepatitis C virus type 1b.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31647;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H05-5;
RA   Ogata S.;
RT   "CORRELATION BETWEEN SECONDARY STRUCTURE OF AN AMINO-TERMINAL PORTION
RT   OF THE NONSTRUCTURAL PROTEIN 3 OF HEPATITIS C VIRUS AND DEVELOPMENT OF
RT   HEPATOCELLULAR CARCINOMA.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB072048; BAB88230.1; -.
DR   InterPro; IPR004109; HCV_NS3.
DR   Pfam; PF02907; HCV_NS3; 1.
FT   NON_TER       1      1
FT   NON_TER     180    180
SQ   SEQUENCE   180 AA;  18913 MW;  895FB23253DA6CD6 CRC64;

  Query Match             97.7%;  Score 43;  DB 12;  Length 180;
  Best Local Similarity   88.9%;  Pred. No. 1.1;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              |||||:|||
Db         13 LLGCIVTSL 21


RESULT 4
Q8QVX3
ID   Q8QVX3      PRELIMINARY;      PRT;   180 AA.
AC   Q8QVX3;
```

```
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Polyprotein (Fragment).
OS   Hepatitis C virus type 1b.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31647;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=82;
RA   Ogata S.;
RT   "CORRELATION BETWEEN SECONDARY STRUCTURE OF AN AMINO-TERMINAL PORTION
RT   OF THE NONSTRUCTURAL PROTEIN 3 OF HEPATITIS C VIRUS AND DEVELOPMENT OF
RT   HEPATOCELLULAR CARCINOMA.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB072102; BAB88284.1; -.
DR   InterPro; IPR004109; HCV_NS3.
DR   Pfam; PF02907; HCV_NS3; 1.
FT   NON_TER       1      1
FT   NON_TER     180    180
SQ   SEQUENCE   180 AA;  18808 MW;  CE520CF0BCA4E1E2 CRC64;

  Query Match             97.7%;  Score 43;  DB 12;  Length 180;
  Best Local Similarity   88.9%;  Pred. No. 1.1;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              |||||:|||
Db         13 LLGCIVTSL 21


RESULT 5
Q8QP94
ID   Q8QP94      PRELIMINARY;      PRT;   403 AA.
AC   Q8QP94;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Genome polyprotein (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=687;
RX   MEDLINE=21904745; PubMed=11907242;
RA   Kalinina O., Norder H., Mukomolov S., Magnius L.O.;
RT   "A natural intergenotypic recombinant of hepatitis C virus identified
RT   in St. Petersburg.";
RL   J. Virol. 76:4034-4043(2002).
DR   EMBL; AY070171; AAL58587.1; -.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     403    403
SQ   SEQUENCE   403 AA;  43868 MW;  EA5C9FACB9348C8A CRC64;

  Query Match             97.7%;  Score 43;  DB 12;  Length 403;
  Best Local Similarity   88.9%;  Pred. No. 2.2;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              |||||:|||
Db        323 LLGCIVTSL 331
```

```
RESULT 6
Q9QIY5
ID   Q9QIY5      PRELIMINARY;      PRT;  3010 AA.
AC   Q9QIY5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD3-2;
RX   MEDLINE=20013325; PubMed=10544098;
RA   Nagayama K., Kurosaki M., Enomoto N., Maekawa S.y., Miyasaka Y.,
RA   Tazawa J.i., Izumi N., Marumo F., Sato C.;
RT   "Time-related changes in full-length hepatitis C virus and hepatitis
RT   activity.";
RL   Virology 263:244-253(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD3-2;
RA   Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,
RA   Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF165050; AAD56185.1; -.
DR   HSSP; P26663; 1JXP.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS50507; RDRP_POSITIVE; 1.
DR   PROSITE; PS50521; RDRP_VIRAL; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327430 MW;  15190E3463DE8C35 CRC64;

  Query Match             97.7%;  Score 43;  DB 12;  Length 3010;
  Best Local Similarity   88.9%;  Pred. No. 12;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              :||||:|||
Db       1039 LLGCIVTSL 1047
```

```
RESULT 7
Q9QIY6
ID   Q9QIY6      PRELIMINARY;      PRT;  3010 AA.
AC   Q9QIY6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD3-1;
RX   MEDLINE=20013325; PubMed=10544098;
RA   Nagayama K., Kurosaki M., Enomoto N., Maekawa S.y., Miyasaka Y.,
RA   Tazawa J.i., Izumi N., Marumo F., Sato C.;
RT   "Time-related changes in full-length hepatitis C virus and hepatitis
RT   activity.";
RL   Virology 263:244-253(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD3-1;
RA   Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,
RA   Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF165049; AAD56184.1; -.
DR   HSSP; P26663; 1JXP.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS50507; RDRP_POSITIVE; 1.
DR   PROSITE; PS50521; RDRP_VIRAL; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327368 MW;  998C7F293EAAEC8D CRC64;

  Query Match             97.7%;  Score 43;  DB 12;  Length 3010;
  Best Local Similarity   88.9%;  Pred. No. 12;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              :||||:|||
Db       1039 LLGCIVTSL 1047
```

```
RESULT 8
Q9QIY2
ID   Q9QIY2      PRELIMINARY;      PRT;  3010 AA.
AC   Q9QIY2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD5-1;
RX   MEDLINE=20013325; PubMed=10544098;
RA   Nagayama K., Kurosaki M., Enomoto N., Maekawa S.y., Miyasaka Y.,
RA   Tazawa J.i., Izumi N., Marumo F., Sato C.;
RT   "Time-related changes in full-length hepatitis C virus and hepatitis
RT   activity.";
RL   Virology 263:244-253(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD5-1;
RA   Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,
RA   Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF165053; AAD56188.1; -.
DR   HSSP; P27958; 1HEI.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS50507; RDRP_POSITIVE; 1.
DR   PROSITE; PS50521; RDRP_VIRAL; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  326595 MW;  D186BA7E92F0B5E8 CRC64;

  Query Match             97.7%;  Score 43;  DB 12;  Length 3010;
  Best Local Similarity   88.9%;  Pred. No. 12;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              |||||:|||
Db       1039 LLGCIVTSL 1047
```

```
RESULT 9
Q8V638
ID   Q8V638      PRELIMINARY;      PRT;  3010 AA.
AC   Q8V638;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Genome polyprotein (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fanning L.J., Itakura J., Nagayama K., Enomoto N.;
RT   "Characteristics of Hepatitis C viral genome associated with disease
RT   progression in a homogeneous patient population.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF313916; AAL55821.1; -.
DR   InterPro; IPR000745; CytC_heme_bind.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
DR   PROSITE; PS50507; RDRP_POSITIVE; 1.
DR   PROSITE; PS50521; RDRP_VIRAL; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT   NON_TER    3010   3010
SQ   SEQUENCE   3010 AA;  327181 MW;  33AAA6C07251C839 CRC64;

  Query Match             97.7%;  Score 43;  DB 12;  Length 3010;
  Best Local Similarity   88.9%;  Pred. No. 12;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              |||||:|||
Db       1039 LLGCIVTSL 1047


RESULT 10
Q81541
ID   Q81541      PRELIMINARY;      PRT;  3010 AA.
AC   Q81541;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
```

```
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JT;
RX   MEDLINE=92295714; PubMed=1318627;
RA   Tanaka T., Kato N., Nakagawa M., Gotsuyama Y., Cho M.J., Nakazawa I.,
RA   Hijikata M., Ishimura Y., Shimotohno K.;
RT   "Molecular cloning of hepatitis C virus genome from a single Japanese
RT   carrier: sequence variation within the same individual and among
RT   infected individuals.";
RL   Virus Res. 23:39-53(1992).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; D11355; BAA18894.1; -.
DR   HSSP; P26663; 1JXP.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS50507; RDRP_POSITIVE; 1.
DR   PROSITE; PS50521; RDRP_VIRAL; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT   CHAIN         1    191       PUT. P22.
FT   CHAIN       192    383       PUT. GP35.
FT   CHAIN       384    729       PUT. GP70.
FT   CHAIN       730   1006       PUT. NS2.
FT   CHAIN      1007   1615       PUT. NS3.
FT   CHAIN      1616   1862       PUT. NS4A.
FT   CHAIN      1863   2013       PUT. NS4B.
FT   CHAIN      2014   3010       PUT. NS5.
SQ   SEQUENCE   3010 AA;  326564 MW;  05F0B2102CF90D9D CRC64;

  Query Match             97.7%;  Score 43;  DB 12;  Length 3010;
  Best Local Similarity   88.9%;  Pred. No. 12;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              |||||:|||
Db       1039 LLGCIVTSL 1047


RESULT 11
Q9DTD6
ID   Q9DTD6      PRELIMINARY;      PRT;  3010 AA.
AC   Q9DTD6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCVT221;
RA   Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA   Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA   Mishiro S.;
RT   "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT   with hepatocellular carcinoma: the 'progression score' revisited.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AB049101; BAB18814.1; -.
DR   HSSP; P26663; 1JXP.
DR   InterPro; IPR000345; CytC_heme_bind.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
DR   PROSITE; PS50507; RDRP_POSITIVE; 1.
DR   PROSITE; PS50521; RDRP_VIRAL; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327108 MW;  DE182D810EF78EE4 CRC64;

  Query Match             97.7%;  Score 43;  DB 12;  Length 3010;
  Best Local Similarity   88.9%;  Pred. No. 12;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              |||||:|||
Db       1039 LLGCIVTSL 1047


RESULT 12
Q913D4
ID   Q913D4      PRELIMINARY;      PRT;  3011 AA.
AC   Q913D4;
```

```
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Guntaka R.V., Munpally S.K., Khaja M.N., Kota K.K., Ramana V.K.,
RA   Swaminathan S., Sakata Y., Habeebullah C.M.;
RT   "Nucleotide Sequence of Indian strain of Hepatitis C Virus.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AY051292; AAK95832.1; -.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR000345; CytC_heme_bind.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
DR   PROSITE; PS50507; RDRP_POSITIVE; 1.
DR   PROSITE; PS50521; RDRP_VIRAL; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3011 AA;  327234 MW;  57A2196B4227B5C CRC64;

  Query Match             97.7%;  Score 43;  DB 12;  Length 3011;
  Best Local Similarity   88.9%;  Pred. No. 12;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              !!!!|:|||
Db       1039 LLGCIVTSL 1047


RESULT 13
Q91RP5
ID   Q91RP5      PRELIMINARY;      PRT;   181 AA.
AC   Q91RP5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   NS3 protease (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pt.3Q;
RA   Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT   "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT   Clinical Strains of the Hepatitis C Virus.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF369258; AAK54583.1; -.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR004109; HCV_NS3.
DR   Pfam; PF02907; HCV_NS3; 1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     181    181
SQ   SEQUENCE   181 AA;  19056 MW;  BAF89690AD2971DB CRC64;

  Query Match             95.5%;  Score 42;  DB 12;  Length 181;
  Best Local Similarity   88.9%;  Pred. No. 1.8;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLGCIITSL 9
              :|||||||:|
Db         13 ILGCIITSL 21


RESULT 14
Q9J3H9
ID   Q9J3H9      PRELIMINARY;      PRT;  3010 AA.
AC   Q9J3H9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD13;
RA   Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT   "Characteristics of hepatitis C viral genome associated with disease
RT   progression.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF207754; AAF65944.1; -.
DR   HSSP; P26663; 1JXP.
DR   InterPro; IPR000345; CytC_heme_bind.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; HCV_NS3.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
```

DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 326984 MW; AF12CC60E0A4B078 CRC64;

Query Match 95.5%; Score 42; DB 12; Length 3010;
Best Local Similarity 88.9%; Pred. No. 19;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLGCIITSL 9
:|:|:||||
Db 1039 MLGCIITSL 1047

RESULT 15
Q9J3G5
ID Q9J3G5 PRELIMINARY; PRT; 3010 AA.
AC Q9J3G5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD27;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
RT progression.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF207768; AAF65958.1; -.
DR HSSP; P26663; 1JXP.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSvir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 2.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 327177 MW; 47A51DD7678DE62F CRC64;

Query Match 95.5%; Score 42; DB 12; Length 3010;
Best Local Similarity 88.9%; Pred. No. 19;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLGCIITSL 9
:|:|:||||
Db 1039 LIGCIITSL 1047

Search completed: September 29, 2003, 19:08:26
Job time : 32 secs